(12)

# **EUROPEAN PATENT APPLICATION**

(43) Date of publication: 12.04.2000 Bulletin 2000/15

(21) Application number: 99119184.2

(22) Date of filing: 07.10.1999

(51) Int CI.7: **C12N 15/12**, C07K 14/75, C07K 14/78, C07K 14/51, C07K 14/47, C07K 14/495, C12P 21/02

(84) Designated Contracting States:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU

MC NL PT SE

Designated Extension States:

AL LT LV MK RO SI

(30) Priority: 09.10.1998 US 169768

(71) Applicant: United States Surgical Corporation Norwalk, Connecticut 06856 (US)

(72) Inventors:

Gruskin, Elliott A.
 Killingworth, CT 06419 (US)

 Buechter, Douglas D. Wallingford, CT 06492 (US)

 Zhang, Guanghui Guilford, CT 06473 (US)

(11)

 Connolly, Kevin Los Angeles, CA 90066 (US)

(74) Representative: HOFFMANN - EITLE
Patent- und Rechtsanwälte
Arabellastrasse 4
81925 München (DE)

(54) Extracellular matrix proteins with modified amino acid

(57) Incorporation of certain amino acid analogs into polypeptides produced by cells which do not ordinarily provide polypeptides containing such amino acid analogs is accomplished by subjecting the cells to growth media containing such amino acid analogs. The degree of incorporation can be regulated by adjusting the concentration of amino acid analogs in the media and/or by adjusting osmolality of the media. Such incorporation allows the chemical and physical characteristics of

polypeptides to be altered and studied. In addition, nucleic acid and corresponding proteins including a domain from a physiologically active peptide and a domain from an extracellular matrix protein which is capable of providing a self-aggregate are provided. Human extracellular matrix proteins capable of providing a self-aggregate collagen are provided which are produced by prokaryotic cells. Preferred codon usage is employed to produce extracellular matrix proteins in prokaryotics.

EP 0 992 586 A2

**EEST AVAILABLE COPY** 

# Description

1.3

5

4

1000

20

25

30

35

45

50

# **BACKGROUND**

### 1. Technical Field

[0001] Engineered polypeptides and chimeric polypeptides having incorporated amino acids which enhance or otherwise modify properties of such polypeptides.

# 2. Description of Related Art

[0002] Genetic engineering allows polypeptide production to be transferred from one organism to another. In doing so, a portion of the production apparatus indigenous to an original host is transplanted into a recipient. Frequently, the original host has evolved certain unique processing pathways in association with polypeptide production which are not contained in or transferred to the recipient. For example, it is well known that mammalian cells incorporate a complex set of post-translational enzyme systems which impart unique characteristics to protein products of the systems. When a gene encoding a protein normally produced by mammalian cells is transferred into a bacterial or yeast cell, the protein may not be subjected to such post translational modification and the protein may not function as originally intended. [0003] Normally, the process of polypeptide or protein synthesis in living cells involves transcription of DNA into RNA and translation of RNA into protein. Three forms of RNA are involved in protein synthesis: messenger RNA (mRNA) carries genetic information to ribosomes made of ribosomal RNA (rRNA) while transfer RNA (tRNA) links to free amino acids in the cell pool. Amino acid/tRNA complexes line up next to codons of mRNA, with actual recognition and binding being mediated by tRNA. Cells can contain up to twenty amino acids which are combined and incorporated in sequences of varying permutations into proteins. Each amino acid is distinguished from the other nineteen amino acids and charged to tRNA by enzymes known as aminoacyl-tRNA synthetases. As a general rule, amino acid/tRNA complexes are quite specific and normally only a molecule with an exact stereochemical configuration is acted upon by a particular aminoacyl-tRNA synthetase.

[0004] In many living cells some amino acids are taken up from the surrounding environment and some are synthesized within the cell from precursors, which in turn have been assimilated from outside the cell. In certain instances, a cell is auxotrophic, i.e., it requires a specific growth substance beyond the minimum required for normal metabolism and reproduction which it must obtain from the surrounding environment. Some auxotrophs depend upon the external environment to supply certain amino acids. This feature allows certain amino acid analogs to be incorporated into proteins produced by auxotrophs by taking advantage of relatively rare exceptions to the above rule regarding stere-ochemical specificity of aminoacyl-tRNA synthetases. For example, proline is such an exception, i.e., the amino acid activating enzymes responsible for the synthesis of prolyl-tRNA complex are not as specific as others. As a consequence certain proline analogs have been incorporated into bacterial, plant, and animal cell systems. See Tan et al., Proline Analogues Inhibit Human Skin Fibroblast Growth and Collagen Production in Culture, Journal of Investigative Dermatology, 80:261-267(1983).

[0005] A method of incorporating unnatural amino acids into proteins is described, e.g., in Noren et al., A General Method For Site-Specific Incorporation of Unnatural Amino Acids Into Proteins, Science, Vol. 244, pp. 182-188 (1989) wherein chemically acylated suppressor tRNA is used to insert an amino acid in response to a stop codon substituted for the codon encoding residue of interest. See also, Dougherty et al., Synthesis of a Genetically Engineered Repetitive Polypeptide Containing Periodic Selenomethionine Residues, Macromolecules, Vol. 26, No. 7, pp. 1779-1781 (1993), which describes subjecting an *E. coli* methionine auxotroph to selenomethionine containing medium and postulates on the basis of experimental data that selenomethionine may completely replace methionine in all proteins produced by the cell.

cis-Hydroxy-L-proline has been used to study its effects on collagen by incorporation into eukaryotic cells such as cultured normal skin fibroblasts (see Tan et al., supra) and tendon cells from chick embryos (see e.g., Uitto et al., Procollagen Polypeptides Containing cis-4-Hydroxy-L-proline are Overglycosylated and Secreted as Nonhelical Pro-γ-Chains, Archives of Biochemistry and Biophysics, 185:1:214-221(1978)). However, investigators found that trans-4-hydroxyproline would not link with proline specific tRNA of prokaryotic E. coli. See Papas et al., Analysis of the Amino Acid Binding to the Proline Transfer Ribonucleic Acid Synthetase of Escherichia coli, Journal of Biological Chemistry, 245:7:1588-1595(1970). Another unsuccessful attempt to incorporate trans-4-hydroxyproline into prokaryotes is described in Deming et al., In Vitro Incorporation of Proline Analogs into Artificial Proteins, Poly. Mater. Sci. Engin. Proceed., Vol. 71, p. 673-674 (1994). Deming et al. report surveying the potential for incorporation of certain proline analogs, i.e., L-azetidine-2-carboxylic acid, L-γ-thiaproline, 3,4-dehydroproline and L-trans-4-hydroxyproline into artificial proteins expressed in E. coli cells. Only L-azetidine-2-carboxylic acid, L-γ-thiaproline and 3,4 dehydroproline are reported as being incorporated into proteins in E. coli cells in vivo.

[0007] Extracellular matrix proteins ("EMPs") are found in spaces around or near cells of multicellular organisms and are typically fibrous proteins of two functional types: mainly structural, e.g., collagen and elastin, and mainly adhesive, e.g., fibronectin and laminin. Collagens are a family of fibrous proteins typically secreted by connective tissue cells. Twenty distinct collagen chains have been identified which assemble to form a total of about ten different collagen molecules. A general discussion of collagen is provided by Alberts, et al., The Cell, Garland Publishing, pp. 802-823 (1989), incorporated herein by reference. Other fibrous or filamentous proteins include Type I IF proteins, e.g., keratins; Type II IF proteins, e.g., vimentin, desmin and glial fibrillary acidic protein; Type III IF proteins, e.g., neurofilament proteins; and Type IV IF proteins, e.g., nuclear laminins.

5

10

15

20

25

30

35

50

[0008] Type I collagen is the most abundant form of the fibrillar, interstitial collagens and is the main component of the extracellular matrix. Collagen monomers consist of about 1000 amino acid residues in a repeating array of Gly-X-Y triplets. Approximately 35% of the X and Y positions are occupied by proline and *trans* 4-hydroxyproline. Collagen monomers associate into triple helices which consist of one  $\alpha 2$  and two  $\alpha 1$  chains. The triple helices associate into fibrils which are oriented into tight bundles. The bundles of collagen fibrils are further organized to form the scaffold for extracellular matrix.

[0009] In mammalian cells, post-translational modification of collagen contributes to its ultimate chemical and physical properties and includes proteolytic digestion of pro-regions, hydroxylation of lysine and proline, and glycosylation of hydroxylated lysine. The proteolytic digestion of collagen involves the cleavage of pro regions from the N and C termini. It is known that hydroxylation of proline is essential for the mechanical properties of collagen. Collagen with low levels of 4-hydroxyproline has poor mechanical properties, as highlighted by the sequelae associated with scurvy. 4-hydroxyproline adds stability to the triple helix through hydrogen bonding and through restricting rotation about C-N bonds in the polypeptide backbone. In the absence of a stable structure, naturally occurring cellular enzymes contribute to degrading the collagen polypeptide.

[0010] The structural attributes of Type I collagen along with its generally perceived biocompatability make it a desirable surgical implant material. Collagen is purified from bovine skin or tendon and used to fashion a variety of medical devices including hemostats, implantable gels, drug delivery vehicles and bone substitutes. However, when implanted into humans bovine collagen can cause acute and delayed immune responses.

[0011] As a consequence, researchers have attempted to produce human recombinant collagen with all of its structural attributes in commercial quantities through genetic engineering. Unfortunately, production of collagen by commercial mass producers of protein such as *E. coli* has not been successful. A major problem is the extensive post-translational modification of collagen by enzymes not present in *E. coli*. Failure of *E. coli* cells to provide proline hydroxylation of unhydroxylated collagen proline prevents manufacture of structurally sound collagen in commercial quantities

[0012] Another problem in attempting to use E. coli to produce human collagen is that E. coli prefer particular codons in the production of polypeptides. Although the genetic code is identical in both prokaryotic and eukaryotic organisms, the particular codon (of the several possible for most amino acids) that is most commonly utilized can vary widely between prokaryotes and eukaryotes. See, Wada, K.-N., Y. Wada, F. Ishibashi, T. Gojobori and T. Ikemura. Nucleic Acids Res. 20, Supplement: 2111-2118, 1992. Efficient expression of heterologous (e.g. mammalian) genes in prokaryotes such as E. coli can be adversely affected by the presence in the gene of codons infrequently used in E. coli and expression levels of the heterologous protein often rise when rare codons are replaced by more common ones. See, e.g., Williams, D.P., D. Regier, D. Akiyoshi, F. Genbauffe and J.R. Murphy. Nucleic Acids Res. 16: 10453-10467, 1988 and Höög, J.-O., H. v. Bahr-Lindström, H. Jörnvall and A. Holmgren, Gene. 43: 13-21, 1986. This phenomenon is thought to be related, at least in part, to the observation that a low frequency of occurrence of a particular codon correlates with a low cellular level of the transfer RNA for that codon. See, Ikemura, T.J. Mol. Biol. 158: 573-597, 1982 and Ikemura, T.J. Mol. Biol. 146: 1-21, 1981. Thus, the cellular tRNA level may limit the rate of translation of the codon and therefore influence the overall translation rate of the full-length protein. See, Ikemura, T.J. Mol. Biol. 146: 1-21, 1981; Bonekamp, F. and F.K. Jensen. Nucleic Acids Res. 16: 3013-3024, 1988; Misra, R. and P. Reeves, Eur. J. Biochem. 152: 151-155, 1985; and Post, L.E., G.D. Strycharz, M. Nomura, H. Lewis and P.P. Lewis. Proc. Natl. Acad. Sci. U.S.A. 76: 1697-1701, 1979. In support of this hypothesis is the observation that the genes for abundant E. coli proteins generally exhibit bias towards commonly used codons that represent highly abundant tRNAs. See, Ikemura, T.J. Mol. Biol. 146: 1-21, 1981; Bonekamp, F. and F.K. Jensen. Nucleic Acids Res. 16: 3013-3024, 1988; Misra, R. and P. Reeves, Eur. J. Biochem. 152: 151-155, 1985; and Post, L.E., G.D. Strycharz, M. Nomura, H. Lewis and P.P. Lewis. Proc. Natl. Acad. Sci. U.S.A. 76: 1697-1701, 1979. In addition to codon frequency, the codon context (i.e. the surrounding nucleotides) can also affect expression.

[0013] Although it would appear that substituting preferred codons for rare codons could be expected to increase expression of heterologous proteins in host organisms, such is not the case. Indeed, "it has not been possible to formulate general and unambiguous rules to predict whether the content of low-usage codons in a specific gene might adversely affect the efficiency of its expression in *E. coli.*" See page 524 of S.C. Makrides (1996), Strategies for Achieving High-Level Expression of Genes in *Escherichia coli.* Microbiological Reviews 60, 512-538. For example, in one

case, various gene fusions between yeast a factor and somatomedin C were made that differed only in coding sequence. In these experiments, no correlation was found between codon bias and expression levels in *E. coli*. Ernst, J.F. and Kawashima, E. (1988), J. Biotechnology, 7, 1-10. In another instance, it was shown that despite the higher frequency of optimal codons in a synthetic β-globin gene compared to the native sequence, no difference was found in the protein expression from these two constructs when they were placed behind the T7 promoter. Hernan et al. (1992), Biochemistry, 31, 8619-8628. Conversely, there are many examples of proteins with a relatively high percentage of rare codons that are well expressed in *E. coli*. A table listing some of these examples and a general discussion can be found in Makoff, A.J. et al. (1989), Nucleic Acids Research, 17, 10191-10202. In one case, introduction of non-optimal, rare arginine codons at the 3' end of a gene actually increased the yield of expressed protein. Gursky, Y.G. and Beabealashvilli, R.Sh. (1994), Gene 148, 15-21.

**[0014]** Failure to provide post-translational modifications such as hydroxylation of proline and the presence in human collagen of rare codons for *E. coli* may be contributing to the difficulties encountered in the expression of human collagen genes in *E. coli*.

# SUMMARY

5

15

20

25

30

35

45

50

[0015] A method of incorporating an amino acid analog into a polypeptide produced by a cell is provided which includes providing a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, providing growth media containing at least one amino acid analog selected from the group consisting of *trans*-4-hydroxyproline, 3-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof and contacting the cell with the growth media wherein the at least one amino acid analog is assimilated into the cell and incorporated into at least one polypeptide.

[0016] Also provided is a method of substituting an amino acid analog of an amino acid in a polypeptide produced by a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, which includes providing a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, providing growth media containing at least one amino acid analog selected from the group consisting of trans-4-hydroxyproline, 3-hydroxyproline, cis-4-fluoro-L-proline and combinations thereof and contacting the cell with the growth media wherein the at least one amino acid analog is assimilated into the cell and incorporated as a substitution for at least one naturally occurring amino acid in at least one polypeptide.

[0017] A method of controlling the amount of an amino acid analog incorporated into a polypeptide is also provided which includes providing at least a first cell selected from the group consisting of prokaryotic cell and eukaryotic cell, providing a first growth media containing a first predetermined amount of at least one amino acid analog selected from the group consisting of *trans*-4-hydroxyproline, 3-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof and contacting the first cell with the first growth media wherein a first amount of amino acid analog is assimilated into the first cell and incorporated into at least one polypeptide. At least a second cell selected from the group consisting of prokaryotic cell and eukaryotic cell, is also provided along with a second growth media containing a second predetermined amount of an amino acid analog selected from the group consisting of *trans*-4-hydroxyproline, 3-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof and the at least second cell is contacted with the second growth media wherein a second amount of amino acid analog is assimilated into the second cell and incorporated into at least one polypeptide.

[0018] Also provided is a method of increasing stability of a recombinant polypeptide produced by a cell which includes providing a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, and providing growth media containing an amino acid analog selected from the group consisting of trans-4-hydroxyproline, 3-hydroxyproline, cis-4-fluoro-L-proline and combinations thereof and contacting the cell with the growth media wherein the amino acid analog is assimilated into the cell and incorporated into a recombinant polypeptide, thereby stabilizing the polypeptide.

[0019] A method of increasing uptake of an amino acid analog into a cell and causing formation of an amino acid analog/tRNA complex is also provided which includes providing a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, providing hypertonic growth media containing amino acid analog selected from the group consisting of *trans*-4-hydroxyproline, 3-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof and contacting the cell with the hypertonic growth media wherein the amino acid analog is assimilated into the cell and incorporated into an amino acid analog/tRNA complex. In any of the other above methods, a hypertonic growth media can optionally be incorporated to increase uptake of an aminoacid analog into a cell.

**[0020]** A composition is provided which includes a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, and hypertonic media including an amino acid analog selected from the group consisting of *trans*-4-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof.

[0021] Also provided is a method of producing an Extracellular Matrix Protein (EMP) or a fragment thereof capable of providing a self-aggregate in a cell which does not ordinarily hydroxylate proline which includes providing a nucleic acid sequence encoding the EMP or fragment thereof which has been optimized for expression in the cell by substitution of codons preferred by the cell for naturally occurring codons not preferred by the cell, incorporating the nucleic acid

sequence into the cell, providing hypertonic growth media containing at least one amino acid selected from the group consisting of *trans-4*-hydroxyproline and 3-hydroxyproline, and contacting the cell with the growth media wherein the at least one amino acid is assimilated into the cell and incorporated into the EMP or fragment thereof.

[0022] Nucleic acid encoding a chimeric protein is provided which includes a domain from a physiologically active peptide and a domain from an extracellular matrix protein (EMP) which is capable of providing a self-aggregate. The nucleic acid may be inserted into a cloning vector which can then be incorporated into a cell.

[0023] Also provided is a chimeric protein including a domain from a physiologically active peptide and a domain from an extracellular matrix protein (EMP) which is capable of providing a self aggregate.

[0024] Also provided is human collagen produced by a prokaryotic cell, the human collagen being capable of providing a self aggregate.

[0025] Also provided is nucleic acid encoding a human Extracellular Matrix Protein (EMP) wherein the codon usage in the nucleic acid sequence reflects preferred codon usage in a prokaryotic cell.

# **BRIEF DESCRIPTION OF THE DRAWINGS**

ξ.

15

20

40

45

50

[0026] Figure 1 is a plasmid map illustrating pMAL-c2.

[0027] Figure 2 is a graphical representation of the concentration of intracellular hydroxyproline based upon concentration of *trans*-4-hydroxyproline in growth culture over time.

[0028] Figure 2A is a graphical representation of the concentration of intracellular hydroxyproline as a function of sodium chloride concentration.

[0029] Figures 3A and 3B depict a DNA sequence encoding human Type 1 ( $\alpha_1$ ) collagen (SEQ. ID. NO. 1).

[0030] Figure 4 is a plasmid map illustrating pHuCol.

[0031] Figure 5 depicts a DNA sequence encoding a fragment of human Type 1 ( $\alpha_1$ ) collagen (SEQ. ID. NO.2.).

[0032] Figure 6 is a plasmid map illustrating pHuCol-Fl.

<sup>25</sup> [0033] Figure 7 depicts a DNA sequence encoding a collagen-like peptide wherein the region coding for gene collagen-like peptide is underlined (SEQ, ID, NO, 3).

[0034] Figure 8 depicts an amino acid sequence of a collagen-like peptide (SEQ. ID. NO. 4).

[0035] Figure 9 is a plasmid map illustrating pCLP.

[0036] Figure 10 depicts a DNA sequence encoding mature bone morphogenic protein (SEQ. ID. NO. 5).

30 [0037] Figure 11 is a plasmid map illustrating pCBC.

[0038] Figure 12 is a graphical representation of the percent incorporation of proline and *trans*-4-hydroxyproline into maltose binding protein under various conditions.

[0039] Figure 13 depicts a collagen I (α1)/BMP-2B chimeric amino acid sequence (SEQ. ID. NO. 6).

[0040] Figure 14A-14C depicts a collagen I (α1)/BMP-2B chimeric nucleotide sequence (SEQ. ID. NO. 7).

35 [0041] Figure 15 depicts a collagen I (α1)/TGF-β<sub>1</sub>amino acid sequence (SEQ. ID. NO. 8).

[0042] Figure 16A-16C depict a collagen I ( $\alpha$ 1)/TGF- $\beta_1$  nucleotide sequence (SEQ. ID. NO. 9). Lower case lettering indicates non-coding sequence.

[0043] Figures 17A-17B depict a collagen I (α1)/decorin amino acid sequence (SEQ. ID. NO. 10).

[0044] Figure 18 depicts a collagen I (α1)/decorin peptide amino acid sequence (SEQ. ID. NO.11).

[0045] Figures 19A-19D depict a collagen I (α1)/decorin nucleotide sequence (SEQ. ID. NO. 12).

[0046] Figures 20A-20C depict a collagen/decorin peptide nucleotide sequence (SEQ. ID. NO. 13). Lower case lettering indicates non-coding sequence.

[0047] Figure 21 depicts a pMal cloning vector and polylinker cloning site.

[0048] Figure 22 depicts a polylinker cloning site contained in the pMal cloning vector of Fig. 21 (SEQ. ID. NO. 14).

[0049] Figure 23 depicts a pMal cloning vector containing a BMP/collagen nucleotide chimeric construct.

[0050] Figure 24 depicts a pMal cloning vector containing a TGF- $\beta_1$ /collagen nucleotide chimeric construct.

[0051] Figure 25 depicts a pMal cloning vector containing a decorin/collagen nucleotide chimeric construct.

[0052] Figure 26 depicts a pMal cloning vector containing a decorin peptide/collagen nucleotide chimeric construct.

[0053] Figure 27A-27E depicts a human collagen Type I ( $\alpha_1$ ) nucleotide sequence (SEQ. ID. NO. 15) and corresponding amino acid sequence (SEQ. ID. NO. 16).

[0054] Figure 28 is a schematic diagram of the construction of the human collagen gene from synthetic oligonucleotides.

[0055] Figure 29 is a schematic depiction of the amino acid sequence of chimeric proteins GST-CoIECoI (SEQ. ID. NO. 17) and GST-D4 (SEQ. ID. NO. 18).

5 [0056] Figure 30 is a Table depicting occurrence of four proline and four glycine codons in the human Collagen Type I (α₁) gene with optimized codon usage (CoIECoI).

[0057] Figure 31 depicts a gel reflecting expression and dependence of expression of GST-D4 on hydroxyproline.

[0058] Figure 32 depicts a gel showing expression of GST-D4 in hypertonic media.

- [0059] Figure 33 is a graph showing circular dichroism spectra of native and denatured D4 in neutral phosphate buffer.
- [0060] Figure 34 depicts a gel representing digestion of D4 with bovine pepsin.

25

- [0061] Figure 35 depicts a gel representing expression of GST-H Col and GST-ColECol under specified conditions.
- [0062] Figure 36 depicts a gel representing expression of GST-CM4 in media with or without Nacl and either proline or hydroxyproline.
  - [0063] Figure 37 depicts a gel of six hour post induction samples of GST-CM4 expressed in *E. coli* with varying concentrations of NaCl.
  - [0064] Figure 38 depicts a gel of 4 hour post induction samples of GST-CM4 expressed in *E. coli* with constant amounts of hydroxyproline and varying amounts of proline.
- [0065] Figures 39A-39E depict the nucleotide (SEQ. ID. NO. 19) and amino acid (SEQ. ID. NO. 20) sequence of HuCol<sup>Ec</sup>, the helical region of human Type I (α<sub>1</sub>) collagen plus 17 amino terminal extra-helical amino acids and 26 carboxy terminal extra-helical amino acids with codon usage optimized for E. coli.
  - [0066] Figure 40 depicts sequence and restriction maps of synthetic oligos used to reconstruct the first 243 base pairs of the human Type I ( $\alpha_1$ ) collagen gene with optimized *E. coli* codon usage. The synthetic oligos are labelled N1-1 (SEQ. ID. NO. 21), N1-2 (SEQ. ID. NO. 22), N1-3 (SEQ. ID. NO. 23) and N1-4 (SEQ. ID. NO. 24).
  - **[0067]** Figure 41 depicts a plasmid map of pBSN1-1 containing a 114 base pair fragment of human collagen Type I  $(\alpha_1)$  with optimized *E. coli* codon usage.
  - [0068] Figure 42 depicts the nucleotide (SEQ. ID. NO. 25) and amino acid (SEQ. ID. NO. 26) sequence of a fragment of human collagen Type I ( $\alpha_1$ ) gene with optimized *E. coli* codon usage encoded by plasmid pBSN1-1.
- [0069] Figure 43 depicts a plasmid map of pBSN1-2 containing a 243 base pair fragment of human collagen Type I  $(\alpha_1)$  with optimized *E. coli* codon usage.
  - [0070] Figure 44 depicts the nucleotide (SEQ. ID. NO. 27) and amino acid (SEQ. ID. NO. 28) sequence of a fragment of human collagen Type I ( $\alpha_1$ ) gene with optimized *E. coli* codon usage encoded by plasmid pBSN1-2.
  - [0071] Figure 45 depicts a plasmid map of pHuCol<sup>Ec</sup> containing human collagen Type I ( $\alpha_1$ ) with optimized E. coli codon usage.
  - [0072] Figure 46 depicts a plasmid map of pTrc N1-2 containing a 234 nucleotide human collagen Type I ( $\alpha_1$ ) fragment with optimized *E. coli* codon usage.
  - [0073] Figure 47 depicts a plasmid map of pN1-3 containing a 360 nucleotide human collagen Type I ( $\alpha_1$ ) fragment with optimized *E. coli* codon usage.
- 30 [0074] Figure 48 depicts a plasmid map of pD4 containing a 657 nucleotide human collagen Type I (α<sub>1</sub>) 3' fragment with optimized *E. coli* codon usage.
  - **[0075]** Figures 49A-49E depict the nucleotide (SEQ. ID. NO. 29) and amino acid (SEQ. ID. NO. 30) sequence of a helical region of human Type I ( $\alpha_2$ ) collagen plus 11 amino terminal extra-helical amino acids and 12 carboxy terminal extrahelical amino acids.
- [0076] Figures 50A-50E depict the nucleotide (SEQ. ID. NO. 31) and amino acid (SEQ. ID. NO. 32) sequence of HuCol( $\alpha_2$ )<sup>Ec</sup>, the helical region of human Type I ( $\alpha_2$ ) collagen plus 11 amino terminal extra-helical amino acids and 12 carboxy terminal extra-helical amino acids with codon usage optimized for *E. coli*.
  - [0077] Figure 51 depicts sequence and restriction maps of synthetic oligos used to reconstruct the first 240 base pairs of human Type I ( $\alpha_2$ ) collagen gene with optimized *E. coli* codon usage. The synthetic oligos are labelled N1-1
  - ( $\alpha$ 2) (SEQ. ID. NO. 33), N1-2 ( $\alpha$ 2) (SEQ. ID. NO. 34), N1-3 ( $\alpha$ 2) (SEQ. ID. NO. 35) and N1-4 ( $\alpha$ 2) (SEQ. ID. NO. 36). **[0078]** Figure 52 depicts a plasmid map of pBsN1-I ( $\alpha$ 2) containing a 117 base pair fragment of human collagen Type I ( $\alpha$ 2) with optimized *E. coli* codon usage.
    - [0079] Figure 53 depicts a plasmid map of pBSN1-2 ( $\alpha_2$ ) containing a 240 base pair fragment of human collagen Type I ( $\alpha_2$ ) with optimized *E. coli* codon usage.
- [0080] Figure 54 depicts the nucleotide (SEQ. ID. NO. 37) and amino acid (SEQ. ID. NO. 38) sequence of a fragment of human collagen Type I ( $\alpha_2$ ) gene with optimized *E. coli* usage encoded by plasmid pBSN1-2( $\alpha_2$ ).
  - [0081] Figure 55 depicts a plasmid map of pHucol( $\alpha_2$ )<sup>Ec</sup> containing the entire human collagen Type I ( $\alpha_2$ ) gene with optimized *E. coli* codon usage.
  - [0082] Figure 56 depicts a plasmid map of pN1-2 ( $\alpha_2$ ) containing a 240 base pair fragment of human collagen Type I( $\alpha_2$ ) with optimized *E. coli* codon usage.
  - [0083] Figure 57 depicts a gel reflecting expression of GST and TGF-β1 under specified conditions.
  - [0084] Figure 58 depicts a gel reflecting expression of MBP, FN-BMP-2A, FN-TGF-β1 and FN under specified conditions.
  - [0085] Figure 59 depicts a gel showing expression of GST-Coll under specified conditions.
- 55 [0086] Figure 60 depicts a plasmid map of pGST-CM4 containing the gene for glutathione S- transferase fused to the gene for collagen mimetic 4.
  - [0087] Figure 61 depicts the nucleotide (SEQ. ID. NO. 39) and amino acid (SEQ. ID. NO. 40) sequence of collagen mimetic 4.

[0088] Figure 62A depicts a chromatogram of the elution of hydroxyproline containing collagen mimetic 4 from a Poros RP2 column. The arrow indicates the peak containing hydroxyproline containing collagen mimetic 4.

[0089] Figure 62B depicts a chromatogram of the elution of proline-containing collagen mimetic 4 from a Poros RP2 column. The arrow indicates the peak containing proline containing collagen mimetic 4.

[0090] Figure 63A depicts a chromatogram of a proline amino acid standard (250 pmol).

5

20

25

35

45

50

55

- [0091] Figure 63B depicts a chromatogram of a hydroxyproline amino acid standard (250 pmol).
- [0092] Figure 63C depicts an amino acid analysis chromatogram of the hydrolysis of proline containing collagen mimetic 4.
- [0093] Figure 63D depicts an amino acid analysis chromatogram of the hydrolysis of hydroxyproline containing collagen mimetic 4.
- [0094] Figure 64 is a graph of OD600 versus time for cultures of *E. coli* JM109 (F-) grown to plateau and then supplemented with various amino acids.
- [0095] Figure 65 depicts a plasmid map of pcEc- $\alpha$ 1 containing the gene for HuCol( $\alpha$ 1)<sup>Ec</sup>.
- [0096] Figure 66 depicts a plasmid map of pcEc- $\alpha$ 2 containing the gene for HuCol( $\alpha$ 2)<sup>Ec</sup>
- [0097] Figure 67 depicts a plasmid map of pD4- $\alpha$ 1 containing the gene for a 219 amino acid C-terminal fragment of Type I ( $\alpha$ 1) human collagen with optimized *E. coli* codon usage fused to the gene for glutathione S-transferase.
- [0098] Figure 68 depicts a plasmid map of pD4- $\alpha$ 2 containing the gene for a 207 amino acid C-terminal fragment of Type I ( $\alpha$ 2) human collagen with optimized *E. coli* codon usage fused to the gene for glutathione S-transferase.
- [0099] Figure 69 depicts the predicted amino acid sequence from the DNA sequence of the first 13 amino acid acids of protein D4-α1 (SEQ. ID. NO. 41) and the amino acid sequence as experimentally determined (SEQ. ID NO. 42).
- [0100] Figure 70 depicts the mass spectrum of hydroxyproline containing D4- $\alpha$ 1.
- [0101] Figure 71 depicts the nucleotide sequence of a 657 nucleotide human collagen Type I (α1)3' fragment with optimized *E. coli* codon usage designated D4 (SEQ. ID. NO. 43).
- [0102] Figure 72 depicts the amino acid sequence of a 219 amino acid C-terminal fragment of human collagen Type I (α1) designed D4 (SEQ. ID. NO. 44).
- [0103] Figure 73 is a plasmid map illustrating pGEX-4T. 1 containing the gene for glutatione S-transferase.
- [0104] Figure 74 is a plasmid map illustrating pTrc-TGF containing the gene for the mature human TGF-β1 polypeptide.
- [0105] Figure 75 is a plasmid map illustrating pTrc-Fn containing the gene for a 70 kDa fragment of human fibronectin.
- 30 [0106] Figure 76 is a plasmid map illustrating pTrc-Fn-TGF containing the gene for a fusion protein of a 70 kDA fragment of human fibronectin and the mature human TGF-β1 polypeptide.
  - [0107] Figure 77 is a plasmid map illustrating pTrc-Fn-BMP containing the gene for a fusion protein of a 70 kDa fragment of human fibronectin and human bone morphogenic protein 2A.
  - [0108] Figure 78 is a plasmid map illustrating pGEX-HuColl<sup>Ec</sup> containing the gene for a fusion between glutathione S-transferase and Type I ( $\alpha$ 1) human collagen with optimized *E. coli* codon usage.
  - [0109] Figure 79 depicts the nucleotide sequence of a 627 nucleotide human collagen Type I ( $\alpha$ 2) 3' fragment with optimized *E. coli* codon usage (SEQ. ID. NO.45).
  - [0110] Figure 80 depicts the amino acid sequence of a 209 amino acid C-terminal fragment of human collagen Type I (α2) (SEQ. ID. NO. 46).
- 40 [0111] Figure 81 depicts the sequence of synthetic oligos used to reconstruct the first 282 base pairs of the gene for the carboxy terminal 219 amino acids of human Type I (α1) collagen with optimized E. coli codon usage designated N4-1 (SEQ. ID. NO. 47), N4-2 (SEQ. ID. NO. 48), N4-3 (SEQ. ID. NO. 49) and N4-4 (SEQ. ID. NO. 50).

# **DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS**

[0112] Prokaryotic cells and eukaryotic cells can unexpectedly be made to assimilate and incorporate *trans*-4-hydroxyproline into proteins contrary to both Papas et al. and Deming et al., supra. Such assimilation and incorporation is especially useful when the structure and function of a polypeptide depends on post translational hydroxylation of proline not provided by the native protein production system of a recombinant host. Thus, prokaryotic bacteria such as *E. coli* and eukaryotic cells such as *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis* and *Schizosaccharomyces pombe* that ordinarily do not hydroxylate proline and additional eukaryotes such as insect cells including lepidopteran cell lines including *Spodoptera fiugiperda*, *Trichoplasia ni*, *Heliothis virescens*, *Bombyx mori* infected with a baculovirus; CHO cells, COS cells and NIH 3T3 cells which fail to adequately produce certain polypeptides whose structure and function depend on such hydroxylation can be made to produce polypeptides having hydroxylated prolines. Incorporation includes adding *trans*-4-hydroxyproline to a polypeptide, for example, by first changing an amino acid to proline, creating a new proline position that can in turn be substituted with *trans*-4-hydroxyproline or substituting a naturally occurring proline in a polypeptide with *trans*-4-hydroxyproline as well.

[0113] The process of producing recombinant polypeptides in mass producing organisms is well known. Replicable

expression vectors such as plasmids, viruses, cosmids and artificial chromosomes are commonly used to transport genes encoding desired proteins from one host to another. It is contemplated that any known method of cloning a gene, ligating the gene into an expression vector and transforming a host cell with such expression vector can be used in furtherance of the present disclosure.

[0114] Not only is incorporation of *trans*-4-hydroxyproline into polypeptides which depend upon *trans*-4-hydroxyproline for chemical and physical properties useful in production systems which do not have the appropriate systems for converting proline to *trans*-4-hydroxyproline, but useful as well in studying the structure and function of polypeptides which do not normally contain *trans*-4-hydroxyproline. It is contemplated that the following amino acid analogs may also be incorporated in accordance with the present disclosure: *trans*-4 hydroxyproline, 3-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof (hereinafter referred to as the "amino acid analogs"). Use of prokaryotes and eukaryotes is desirable since they allow relatively inexpensive mass production of such polypeptides. It is contemplated that the amino acid analogs can be incorporated into any desired polypeptide. In a preferred embodiment the prokaryotic cells and eukaryotic cells are starved for proline by decreasing or eliminating the amount of proline in growth media prior to addition of an amino acid analog herein.

[0115] Expression vectors containing the gene for maltose binding protein (MBP), e.g., see Figure 1 illustrating plasmid pMAL-c2, commercially available from New England Bio-Labs, are transformed into prokaryotes such as *E. coli* proline auxotrophs or eukaryotes such as *S. cerevisiae* auxotrophs which depend upon externally supplied proline for protein synthesis and anabolism. Other preferred expression vectors for use in prokaryotes are commercially available plasmids which include pKK-223 (Pharmacia), pTRC (Invitrogen), pGEX (Pharmacia), pET (Novagen) and pQE (Quiagen). It should be understood that any suitable expression vector may be utilized by those with skill in the art.

15

25

30

35

45

55

[0116] Substitution of the amino acid analogs for proline in protein synthesis occurs since prolyl tRNA synthetase is sufficiently promiscuous to allow misacylation of proline tRNA with any one of the amino acid analogs. A sufficient quantity, i.e., typically ranging from about .001M to about 1.0 M, but more preferably from about .005M to about 0.5M of the amino acid analog(s) is added to the growth medium for the transformed cells to compete with proline in cellular uptake. After sufficient time, generally from about 3Q minutes to about 24 hours or more, the amino acid analog(s) is assimilated by the cell and incorporated into protein synthetic pathways. As can be seen from Figures 2 and 2A, intracellular concentration of trans-4-hydroxyproline increases by increasing the concentration of sodium chloride in the growth media. In a preferred embodiment the prokaryotic cells and/or eukaryotic cells are starved for proline by decreasing or eliminating the amount of proline in growth media prior to addition of an amino acid analog herein.

[0117] Expression vectors containing the gene for human Type I (α1) collagen (DNA sequence illustrated in Figures 3 and 3A; plasmid map illustrated in Figure 4) are transformed into prokaryotic or eukaryotic proline auxotrophs which depend upon externally supplied proline for protein synthesis and anabolism. As above, substitution of the amino acid analog(s) occurs since prolyl tRNA synthetase is sufficiently promiscuous to allow misacylation of proline tRNA with the amino acid analog(s). The quantity of amino acid analog(s) in media given above is again applicable.

[0118] Expression vectors containing DNA encoding fragments of human Type 1 ( $\alpha$ 1) collagen (e.g., DNA sequence illustrated in Figure 5 and plasmid map illustrated in Figure 6) are transformed into prokaryotic or eukaryotic auxotrophs as above. Likewise, expression vectors containing DNA encoding collagen-like polypeptide (e.g., DNA sequence illustrated in Figure 7, amino acid sequence illustration in Figure 8 and plasmid map illustrated in Figure 9) can be used to transform prokaryotic or eukaryotic auxotrophs as above. Collagen-like peptides are those which contain at least partial homology with collagen and exhibit similar chemical and physical characteristics to collagen. Thus, collagen-like peptides consist, e.g., of repeating arrays of Gly-X-Y triplets in which about 35% of the X and Y positions are occupied by proline and 4-hydroxyproline. Collagen-like peptides are interchangeably referred to herein as collagen-like proteins, collagen-like polypeptides, collagen mimetic polypeptides and collagen mimetic. Certain preferred collagen fragments and collagen-like peptides in accordance herewith are capable of assembling into an extracellular matrix. In both collagen fragments and collagen-like peptides as described above, substitution with amino acid analog(s) occurs since prolyl tRNA synthetase is sufficiently promiscuous to allow misacylation of proline tRNA with one or more of the amino acid analog(s). The quantity of amino acid analog(s) given above is again applicable.

[0119] It is contemplated that any polypeptide having an extracellular matrix protein domain such as a collagen, collagen fragment or collagen-like peptide domain can be made to incorporate amino acid analog(s) in accordance with the disclosure herein. Such polypeptides include collagen, a collagen fragment or collagen-like peptide domain and a domain having a region incorporating one or more physiologically active agents such as glycoproteins, proteins, peptides and proteoglycans. As used herein, physiologically active agents exert control over or modify existing physiologic functions in living things. Physiologically active agents include hormones, growth factors, enzymes, ligands and receptors. Many active domains of physiologically active agents have been defined and isolated. It is contemplated that polypeptides having a collagen, collagen fragment or collagen-like peptide domain can also have a domain incorporating one or more physiologically active domains which are active fragments of such physiologically active agents. As used herein, physiologically active agent is meant to include entire peptides, polypeptides, proteins, glycoproteins, proteoglycans and active fragments of any of them. Thus, chimeric proteins are made to incorporate amino acid analog

(s) by transforming a prokaryotic proline auxotroph or a eukaryotic proline auxotroph with an appropriate expression vector and contacting the transformed auxotroph with growth media containing at least one of the amino acid analogs. For example, a chimeric collagen/bone morphogenic protein (BMP) construct or various chimeric collagen/growth factor constructs are useful in accordance herein. Such growth factors are well-known and include insulin-like growth factor, transforming growth factor, platelet derived growth factor and the like. Figure 10 illustrates DNA of BMP which can be fused to the 3' terminus of DNA encoding collagen, DNA encoding a collagen fragment or DNA encoding a collagen-like peptide. Figure 11 illustrates a map of plasmid pCBC containing a collagen/BMP construct. In a preferred embodiment, proteins having a collagen, collagen fragment or collagen-like peptide domain assemble or aggregate to form an extracellular matrix which can be used as a surgical implant. The property of self-aggregation as used herein includes the ability to form an aggregate with the same or similar molecules or to form an aggregate with different molecules that share the property of aggregation to form, e.g., a double or triple helix. An example of such aggregation is the structure of assembled collagen matrices.

[0120] Indeed, chimeric polypeptides which may also be referred to herein as chimeric proteins provide an integrated combination of a therapeutically active domain from a physiologically active agent and one or more EMP moieties. The EMP domain provides an integral vehicle for delivery of the therapeutically active moiety to a target site. The two domains are linked covalently by one or more peptide bonds contained in a linker region. As used herein, integrated or integral means characteristics which result from the covalent association of one or more domains of the chimeric proteins. The therapeutically active moieties disclosed herein are typically made of amino acids linked to form peptides, polypeptides, proteins, glycoproteins or proteoglycans. As used herein, peptide encompasses polypeptides and proteins

15

20

25

30

35

45

[0121] The inherent characteristics of EMPs are ideal for use as a vehicle for the therapeutic moiety. One such characteristic is the ability of the EMPs to form the self-aggregate. Examples of suitable EMPs are collagen, elastin, fibronectin, fibrinogen and fibrin. Fibrillar collagens (Type I, II and III) assemble into ordered polymers and often aggregate into larger bundles. Type IV collagen assembles into sheetlike meshworks. Elastin molecules form filaments and sheets in which the elastin molecules are highly cross-linked to one another to provide good elasticity and high tensile strength. The cross-linked, random-coiled structure of the fiber network allows it to stretch and recoil like a rubber band. Fibronectin is a large fibril forming glycoprotein, which, in one of its forms, consists of highly insoluble fibrils cross-linked to each other by disulfide bonds. Fibrin is an insoluble protein formed from fibrinogen by the proteolytic activity of thrombin during the normal clotting of blood.

[0122] The molecular and macromolecular morphology of the above EMPs defines networks or matrices to provide substratum or scaffolding in integral covalent association with the therapeutically active moiety. The networks or matrices formed by the EMP domain provide an environment particularly well suited for ingrowth of autologous cells involved in growth, repair and replacement of existing tissue. The integral therapeutically active moieties covalently bound within the networks or matrices provide maximum exposure of the active agents to their targets to elicit a desired response.

[0123] Implants formed of or from the present chimeric proteins provide sustained release activity in or at a desired locus or target site. Since it is linked to an EMP domain, the therapeutically active domain of the present chimeric protein is not free to separately diffuse or otherwise be transported away from the vehicle which carries it, absent cleavage of peptide bonds. Consequently, chimeric proteins herein provide an effective anchor for therapeutic activity which allows the activity to be confined to a target location for a prolonged duration. Because the supply of therapeutically active agent does not have to be replenished as often when compared to non-sustained release dosage forms, smaller amounts of therapeutically active agent may be used over the course of therapy. Consequently, certain advantages provided by the present chimeric proteins are a decrease or elimination of local and systemic side effects, less potentiation or reduction in therapeutic activity with chronic use, and minimization of drug accumulation in body tissue with chronic dosing.

[0124] Use of recombinant technology allows manufacturing of non-immunogenic chimeric proteins. The DNA encoding both the therapeutically active moiety and the EMP moiety should preferably be derived from the same species as the patient being treated to avoid an immunogenic reaction. For example, if the patient is human, the therapeutically active moiety as well as the EMP moiety is preferably derived from human DNA.

[0125] Osteogenic/EMP chimeric proteins provide biodegradable and biocompatible agents for inducing bone formation at a desired site. As stated above, in one embodiment, a BMP moiety is covalently linked with an EMP to form chimeric protein. The BMP moiety induces osteogenesis and the extracellular matrix protein moiety provides an integral substratum or scaffolding for the BMP moiety and cells which are involved in reconstruction and growth. Compositions containing the BMP/EMP chimeric protein provide effective sustained release delivery of the BMP moiety to desired target sites. The method of manufacturing such an osteogenic agent is efficient because the need for extra time consuming steps as purifying EMP and then admixing it with the purified BMP are eliminated. An added advantage of the BMP/EMP chimeric protein results from the stability created by the covalent bond between BMP and the EMP, i.e., the BMP portion is not free to separately diffuse away from the EMP, thus providing a more stable therapeutic agent.

[0126] Bone morphogenic proteins are class identified as BMP-1 through BMP-9. A preferred osteogenic protein for use in human patients is human BMP-2B. A BMP-2B/collagen IA chimeric protein is illustrated in Fig. 13 (SEQ. ID. NO. 6). The protein sequence illustrated in Fig. 15 (SEQ. ID. NO. 8) includes a collagen helical domain depicted at amino acids 1-1057 and a mature form of BMP-2B at amino acids 1060-1169. The physical properties of the chimeric protein are dominated in part by the EMP component. In the case of a collagen moiety, a concentrated solution of chimeric protein will have a gelatinous consistency that allows easy handling by the medical practitioner. The EMP moiety acts as a sequestering agent to prevent rapid desorption of the BMP moiety from the desired site and to provide sustained release of BMP activity. As a result, the BMP moiety remains at the desired site and provides sustained release of BMP activity at the desired site for a period of time necessary to effectively induce bone formation. The EMP moiety also provides a matrix which allows a patient's autologous cells, e.g., chondrocytes and the like, which are normally involved in osteogenesis to collect therein and form an autologous network for new tissue growth. The gelatinous consistency of the chimeric protein also provides a useful and convenient therapeutic manner for immobilizing active BMP on a suitable vehicle or implant for delivering the BMP moiety to a site where bone growth is desired.

5

15

20

25

30

45

50

[0127] The BMP moiety and the EMP moiety are optionally linked together by linker sequences of amino acids. Examples of linker sequences used are illustrated within the sequence depicted in Figs. 14A-14C (SEQ. ID. NO. 7), 16A-16C (SEQ. ID. NO. 9), 19A-19C (SEQ. ID. NO. 12) and 20A-20C (SEQ. ID. NO. 13), and are described in more detail below. Linker sequences may be chosen based on particular properties which they impart to the chimeric protein. For example, amino acid sequences such as Ile-Glu-Gly-Arg and Leu-Val-Pro-Arg are cleaved by factor XA and thrombin enzymes, respectively. Incorporating sequences which are cleaved by proteolytic enzymes into chimeric proteins herein provides cleavage at the linker site upon exposure to the appropriate enzyme and separation of the two domains into separate entities. It is contemplated that numerous linker sequences can be incorporated into any of the chimeric proteins.

[0128] In another embodiment, a chimeric DNA construct includes a gene encoding an osteogenic protein or a fragment thereof linked to gene encoding an EMP or a fragment thereof. The gene sequence for various BMPs are known, see, e.g., U.S. Patent Nos. 4,294,753, 4,761,471, 5,106,748, 5,187,076, 5,141,905, 5,108,922, 5,116,738 and 5,168,050, each incorporated herein by reference. A BMP-2B gene for use herein is synthesized by ligating oligonucleotides encoding a BMP protein. The oligonucleotides encoding BMP-2B are synthesized using an automated DNA synthesizer (Beckmen Oligo-1000). In preferred embodiment, the nucleotide sequence encoding the BMP is maximized for expression in *E. coli*. This is accomplished by using *E.coli* utilization tables to translate the sequence of amino acids of the BMP into codons that are utilized most often by *E. coli*. Alternatively, native DNA encoding BMP isolated from mammals including humans may be purified and used.

[0129] The BMP gene and the DNA sequence encoding an extracellular matrix protein are cloned by standard genetic engineering methods as described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor 1989, hereby incorporated by reference.

[0130] The DNA sequence corresponding to the helical and telepeptide region of collagen  $I(\alpha 1)$  is cloned from a human fibroblast cell line. Two sets of polymerase chain reactions are carried out using cDNA prepared by standard methods from AG02261A cells. The first pair of PCR primers include a 5' primer bearing an XmnI linker sequence and a 3' primer bearing the BsmI site at nucleotide number 1722. The resulting PCR product consists of sequence from position 1 to 1722. The second pair of primers includes the BsmI site at 1722 and a linker sequence at the 3' end bearing a Bg1II site. The resulting PCR product consists of sequence from position 1722 to 3196. The complete sequence is assembled by standard cloning techniques. The two PCR products are ligated together at the BsmI site, and the combined clone is inserted into any vector with XmnI-Bg1II sites such as pMAL-c2 vector.

[0131] To clone the BMP-2B gene, total cellular RNA is isolated from human osteosarcoma cells (U-20S) by the method described by Robert E. Farrel Jr. (Academic-Press, CA, 1993 pp. 68-69) (herein incorporated by reference). The integrity of the RNA is verified by spectrophotometric analysis and electrophoresis through agarose gels. Typical yields of total RNA are 50 μg from a 100mm confluent tissue culture dish. The RNA is used to generate cDNA by reverse transcription using the Superscript pre-amplification system by Gibco BRL. The cDNA is used as template for PCR amplification using upstream and downstream primers specific for BMP-2B (GenBank HUMBMP2B accession #M22490). The resulting PCR product consists of BMP-2B sequence from position 1289-1619. The PCR product is resolved by electrophoresis through agarose gels, purified with gene clean (BIO 101) and ligated into pMal-c2 vector (New England Biolabs). The domain of human collagen I(α1) chain is cloned in a similar manner. However, the total cellular RNA is isolated from a human fibroblast cell line (AG02261A human skin fibroblasts).

[0132] A chimeric BMP/EMP DNA construct is obtained by ligating a synthetic BMP gene to a DNA sequence encoding an EMP such as collagen, fibrinogen, fibrin, fibronectin, elastin or laminin. However, chimeric polypeptides herein are not limited to these particular proteins. Figs. 14A-14C (SEQ. ID. NO. 7) illustrate a DNA construct which encodes a BMP-2B/collagen I(al) chimeric protein. The coding sequence for an EMP may be ligated upstream and/or downstream and in-frame with a coding sequence for the BMP. The DNA encoding an EMP may be a portion of the gene or an entire EMP gene. Furthermore, two different EMPs may be ligated upstream and downstream from the BMP.

[0133] The BMP-2B/collagen I(al) chimeric protein illustrated in Figs. 14A-14C includes an XmnI linker sequence at base pairs (bp) 1-19, a collagen domain (bp 20-3190), a BgIII/BamHI linker sequence (bp 3191-3196), a mature form of BMP2b (bp 3197-3529) and a HindIII linker sequence (bp 3530-3535).

[0134] Any combination of growth factor and matrix protein sequences are contemplated including repeating units, or multiple arrays of each segment in any order.

[0135] Incorporation of fragments of both matrix and growth factor proteins is also contemplated. For example, in the case of collagen, only the helical domain may be included. Other matrix proteins have defined domains, such as laminin, which has EGF-like domains. In these cases, specific functionalities can be chosen to achieve desired effects. Moreover, it may be useful to combine domains from disparate matrix proteins, such as the helical region of collagen and the cell attachment regions of fibronectin. In the case of growth factors, specific segments have been shown to be removed from the mature protein by post translational processing. Chimeric proteins can be designed to include only the mature biologically active region. For example, in the case of BMP-2B only the final 110 amino acids are found in the active protein.

10

15

20

30

35

40

or downstream of alternate moieties.

[0136] In another embodiment, a transforming growth factor (TGF) moiety is covalently linked with an EMP to form a chimeric protein. The TGF moiety increases efficacy of the body's normal soft tissue repair response and also induces osteogenesis. Consequently, TGF/EMP chimeric proteins may be used for either or both functions. One of the fundamental properties of the TGF- $\beta$ s is their ability to turn on various activities that result in the synthesis of new connective tissue. See, Piez and Sporn eds., Transforming Growth Factor- $\beta$ s Chemistry, Biology and Therapeutics, Annals of the New York Academy of Sciences, Vol. 593, (1990). TGF- $\beta$  is known to exist in at least five different isoforms. The DNA sequence for Human TGF- $\beta_1$  is known and has been cloned. See Derynck et al., Human Transforming Growth Factor-Beta cDNA Sequence and Expression in Tumour Cell Lines, Nature, Vol. 316, pp. 701-705 (1985), herein incorporated by reference. TGF- $\beta_2$  has been isolated from bovine bone, human glioblastoma cells and porcine platelets. TGF- $\beta_3$  has also been cloned. See ten Dijke, et al., Identification of a New Member of the Transforming Growth Factor- $\beta$  Gene Family, Proc. Natl. Acad. Sci. (USA), Vol. 85, pp. 4715-4719 (1988) herein incorporated by reference.

[0137] A TGF-β/EMP chimeric protein incorporates the known activities of TGF-βs and provides integral scaffolding or substratum of the EMP as described above to yield a composition which further provides sustained release focal delivery at target sites.

[0138] The TGF-β moiety and the EMP moiety are optionally linked together by linker sequences of amino acids. Linker sequences may be chosen based upon particular properties which they impart to the chimeric protein. For example, amino acid sequences such as Ile-Glu-Glyn-Arg and Leu-Val-Pro-Arg are cleaved by Factor XA and Thrombin enzymes, respectively. Incorporating sequences which are cleaved by proteolytic enzymes into the chimeric protein provides cleavage at the linker site upon exposure to the appropriate enzyme and separation of the domains into separate entities. Fig. 15 depicts an amino acid sequence for a TGF-β<sub>1</sub>/collagen IA chimeric protein (SEQ. ID. NO. 8). The illustrated amino acid sequence includes the collagen domain (1-1057) and a mature form of TGF $\beta_1$  (1060-1171). [0139] A chimeric DNA construct includes a gene encoding TGF-β<sub>1</sub> or a fragment thereof, or a gene encoding TGF- $\beta_2$  or a fragment thereof, or a gene encoding TGF- $\beta_3$  or a fragment thereof, ligated to a DNA sequence encoding an EMP protein such as collagen (I-IV), fibrin, fibrinogen, fibronectin, elastin or laminin. A preferred chimeric DNA construct combines DNA encoding TGF-β<sub>1</sub>, a DNA linker sequence, and DNA encoding collagen IA. A chimeric DNA construct containing TGF- $\beta_1$  gene and a collagen I( $\alpha$ 1) gene is shown in Figs. 16A-16C (SEQ. ID. NO. 9). The illustrated construct includes an Xmnl linker sequence (bp 1-19), DNA encoding a collagen domain (bp 20-3190), a Bglll linker sequence (bp 3191-3196), DNA encoding a mature form of TGF-β<sub>1</sub> (3197-3535), and an Xbal linker sequence (bp 3536-3541). [0140] The coding sequence for EMP may be ligated upstream and/or downstream and in-frame with a coding sequence for the TGFB. The DNA encoding the extracellular matrix protein may encode a portion of a fragment of the EMP or may encode the entire EMP. Likewise, the DNA encoding the TGF-β may be one or more fragments thereof or the entire gene. Furthermore, two or more different TGF-\(\beta\)s or two or more different EMPs may be ligated upstream

[0141] In yet another embodiment, a dermatan sulfate proteoglycan moiety, also known as decorin or proteoglycan II, is covalently linked with an EMP to form a chimeric protein. Decorin is known to bind to type I collagen and thus affect fibril formation, and to inhibit the cell attachment-promoting activity of collagen and fibrinogen by binding to such molecules near their cell binding sites. Chimeric proteins which contain a decorin moiety act to reduce scarring of healing tissue. The primary structure of the core protein of decorin has been deduced from cloned cDNA. See Krusius et al., Primary Structure of an Extracellular Matrix Proteoglycan Core Protein-Deduced from Cloned cDNA, Proc. Natl. Acad. Sci. (USA), Vol. 83, pp. 7683-7687 (1986) incorporated herein by reference.

[0142] A decorin/EMP chimeric protein incorporates the known activities of decorin and provides integral scaffolding or substratum of the EMP as described above to yield a composition which allows sustained release focal delivery to target sites. Figs. 17A-17B illustrate a decorin/collagen IA chimeric protein (SEQ. ID. NO. 10) in which the collagen domain includes amino acids 1-1057 and the decorin mature protein incudes amino acids 1060-1388. Fig. 18 illustrates a decorin peptide/collagen IA chimeric protein (SEQ. ID. NO. 11) in which the collagen helical domain includes amino

acids 1-1057 and the decorin peptide fragment includes amino acids 1060-1107. The decorin peptide fragment is composed of P46 to G93 of the mature form of decorin.

[0143] Further provided is a chimeric DNA construct which includes a gene encoding decorin or one or more fragments thereof, optionally ligated via a DNA linker sequence to a DNA sequence encoding an EMP such as collagen (I-IV), fibrin, fibrinogen, fibronectin, elastin or laminin. A preferred chimeric DNA construct combines DNA encoding decorin, a DNA linker sequence, and DNA encoding collagen I(α1). A chimeric DNA construct containing a decorin gene and a collagen I(α1) gene is shown in Figs. 19A-19D (SEQ. ID. NO. 12). The illustrated construct includes an XmnI linker sequence (bp 1-19), DNA encoding a collagen domain (bp 20-3190), a Bg1II linker sequence (bp 3191-3196), DNA encoding a mature form of decorin (bp 3197-4186) and a PstI linker sequence. A chimeric DNA construct containing a decorin peptide gene and a collagen I(α1) gene is shown in Figs. 20A-20C (SEQ. ID. NO. 13). The illustrated construct includes an XmnI linker sequence (bp 1-19), DNA encoding a collagen domain (bp 20-3190), a BgIII linker sequence (bp 3191-3196), DNA encoding a peptide fragment of decorin (bp 3197-3343), and a PstI linker sequence (bp 3344-3349).

[0144] The coding sequence for an EMP may be ligated upstream and/or downstream and in-frame with a coding sequence for decorin. The DNA encoding the EMP may encode a portion or fragment of the EMP or may encode the entire EMP. Likewise, the DNA encoding decorin may be a fragment thereof or the entire gene. Furthermore, two or more different EMPs may be ligated upstream and/or downstream from the DNA encoding decorin moiety.

15

25

30

45

[0145] Any of the above described chimeric DNA constructs may be incorporated into a suitable cloning vector. Fig. 21 depicts a pMal cloning vector containing a polylinker cloning site. Examples of cloning vectors are the plasmids pMal-p2 and pMal-c2 (commercially available from New England Biolabs). The desired chimeric DNA construct is incorporated into a polylinker sequence of the plasmid which contains certain useful restriction endonuclease sites which are depicted in Fig. 22 (SEQ. ID. NO. 14). The pMal-p2 polylinker sequence has XmnI, EcoRI, BamHI, HindIII, XbaI, Sa1I and PstI restriction endonuclease sites which are depicted in Fig 22. The polylinker sequence is digested with an appropriate restriction endonuclease and the chimeric construct is incorporated into the cloning vector by ligating it to the DNA sequences of the plasmid. The chimeric DNA construct may be joined to the plasmid by digesting the ends of the DNA construct and the plasmid with the same restriction endonuclease to generate "sticky ends" having 5' phosphate and 3' hydroxyl groups which allow the DNA construct to anneal to the cloning vector. Gaps between the inserted DNA construct and the plasmid are then sealed with DNA ligase. Other techniques for incorporating the DNA construct into plasmid DNA include blunt end ligation, poly(dA.dT) tailing techniques, and the use of chemically synthesized linkers. An alternative method for introducing the chimeric DNA construct into a cloning vector is to incorporate the DNA encoding the extracellular matrix protein into a cloning vector already containing a gene encoding a therapeutically active moiety.

[0146] The cloning sites in the above-identified polylinker site allow the cDNA for the collagen  $I(\alpha 1)$ /BMP-2B chimeric protein illustrated in Figs. 14A-14C (SEQ. ID. NO. 7) to be inserted between the XmnI and the HindIII sites. The cDNA encoding the collagen  $I(\alpha 1)$ /TGF- $\beta_1$  protein illustrated in Figs. 16A-16C (SEQ. ID. NO. 9) is inserted between the XmnI and the XbaI sites. The cDNA encoding the collagen  $I(\alpha 1)$ /decorin protein illustrated in Figs. 19A-19D (SEQ. ID. NO. 12) inserted between the XmnI and the PstI sites. The cDNA encoding the collagen  $I(\alpha 1)$ /decorin peptide illustrated in Figs. 20A-20C (SEQ. ID. NO. 13) is inserted between the XmnI and PstI sites.

[0147] Plasmids containing the chimeric DNA construct are identified by standard techniques such as gel electrophoresis. Procedures and materials for preparation of recombinant vectors, transformation of host cells with the vectors, and host cell expression of polypeptides are described in Sambrook et al., Molecular Cloning: A Laboratory Manual, supra. Generally, prokaryotic or eukaryotic host cells may be transformed with the recombinant DNA plasmids. Transformed host cells may be located through phenotypic selection genes of the cloning vector which provide resistance to a particular antibiotic when the host cells are grown in a culture medium containing that antibiotic.

[0148] Transformed host cells are isolated and cultured to promote expression of the chimeric protein. The chimeric protein may then be isolated from the culture medium and purified by various methods such as dialysis, density gradient centrifugation, liquid column chromatography, isoelectric precipitation, solvent fractionation, and electrophoresis. However, purification of the chimeric protein by affinity chromatography is preferred whereby the chimeric protein is purified by ligating it to a binding protein and contacting it with a ligand or substrate to which the binding protein has a specific affinity.

[0149] In order to obtain more effective expression of mammalian or human eukaryotic genes in bacteria (prokaryotes), the mammalian or human gene may be placed under the control of a bacterial promoter. A protein fusion and purification system is employed to obtain the chimeric protein. Preferably, any of the above-described chimeric DNA constructs is cloned into a pMal vector at a site in the vector's polylinker sequence. As a result, the chimeric DNA construct is operably fused with the malE gene of the pMal vector. The malE gene encodes maltose binding protein (MBP). Fig. 23 depicts a pMal cloning vector containing a BMP/collagen DNA construct. A spacer sequence coding for 10 asparagine residues is located between the malE sequence and the polylinker sequence. This spacer sequence insulates MBP from the protein of interest. Figs. 24, 25 and 26 depict pMal cloning vectors containing DNA encoding

collagen chimeras with TGF- $\beta_1$ , decorin and a decorin peptide, respectively. The pMal vector containing any of the chimeric DNA constructs fused to the malE gene is transformed into *E. coli*.

5

20

25

30

35

40

45

[0150] The *E. coli* is cultured in a medium which induces the bacteria to produce the maltose-binding protein fused to the chimeric protein. This technique utilizes the P<sub>tac</sub> promoter of the pMal vector. The MBP contains a 26 amino acid N-terminal signal sequence which directs the MBP-chimeric protein through the *E. coli* cytoplasmic membrane. The protein can then be purified from the periplasm. Alternatively, the pMal-c2 cloning vector can be used with this protein fusion and purification system. The pMal-c2 vector contains an exact deletion of the malE signal sequence which results in cytoplasmic expression of the fusion protein. A crude cell extract containing the fusion protein is prepared and poured over a column of amylose resin. Since MBP has an affinity for the amylose it binds to the resin. Alternatively, the column can include any substrate for which MBP has a specific affinity. Unwanted proteins present in the crude extract are washed through the column. The MBP fused to the chimeric protein is eluted from the column with a neutral buffer containing maltose or other dilute solution of a desorbing agent for displacing the hybrid polypeptide. The purified MBP-chimeric protein is cleaved with a protease such as factor Xa protease to cleave the MBP from the chimeric protein. The pMal-p2 plasmid has a sequence encoding the recognition site for protease factor Xa which cleaves after the amino acid sequence Isoleucine-Glutamic acid-Glycine-Arginine of the polylinker sequence.

[0151] The chimeric protein is then separated from the cleaved MBP by passing the mixture over an amylose column. An alternative method for separating the MBP from the chimeric protein is by ion exchange chromatography. This system yields up to 100mg of MBP-chimeric protein per liter of culture. See Riggs, P., in Ausebel, F.M., Kingston, R. E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (eds.) Current Protocols in Molecular Biology, Supplement 19 (16.6.1-16.6.10) (1990) Green Associates/Wiley Interscience, New York, New England Biolabs (cat # 800-65S 9pMALc2) pMal protein fusion and purification system hereby incorporated herein by reference. (See also European Patent No. 286 239 herein incorporated by reference which discloses a similar method for production and purification of a protein such as collagen.)

[0152] Other protein fusion and purification systems may be employed to produce chimeric proteins. Prokaryotes such as *E. coli* are the preferred host cells for expression of the chimeric protein. However, systems which utilize eukaryote host cell lines are also acceptable such as yeast, human, mouse, rat, hamster, monkey, amphibian, insect, algae, and plant cell lines. For example, HeLa (human epithelial), 3T3 (mouse fibroblast), CHO (Chinese hamster ovary), and SP 2 (mouse plasma cell) are acceptable cell lines. The particular host cells that are chosen should be compatible with the particular cloning vector that is chosen.

[0153] Another acceptable protein expression system is the Baculovirus Expression System manufactured by Invitrogen of San Diego, California. Baculoviruses form prominent crystal occlusions within the nuclei of cells they infect. Each crystal occlusion consists of numerous virus particles enveloped in a protein called polyhedrin. In the baculovirus expression system, the native gene encoding polyhedrin is substituted with a DNA construct encoding a protein or peptide having a desired activity. The virus then produces large amounts of protein encoded by the foreign DNA construct. The preferred cloning vector for use with this system is pBlueBac III (obtained from Invitrogen of San Diego, California). The baculovirus system utilizes the *Autograph californica* multiple nuclear polyhidrosis virus (ACMNPV) regulated polyhedrin promoter to drive expression of foreign genes. The chimeric gene, i.e., the DNA construct encoding the chimeric protein, is inserted into the pBlueBac III vector immediately downstream from the baculovirus polyhedrin promoter.

[0154] The pBlueBac III transfer vector contains a B-galactosidase reporter gene which allows for identification of recombinant virus. The B-galactosidase gene is driven by the baculovirus ETL promoter (P<sub>ETL</sub>) which is positioned in opposite orientation to the polyhedrin promoter (P<sub>PH</sub>) and the multiple cloning site of the vector. Therefore, recombinant virus coexpresses B-galactosidase and the chimeric gene.

[0155] Spodoptera frugiperda (Sf9) insect cells are then cotransfected with wild type viral DNA and the pBlueBac III vector containing the chimeric gene. Recombination sequences in the pBlueBac III vector direct the vector's integration into the genome of the wild type baculovirus. Homologous recombination occurs resulting in replacement of the native polyhedrin gene of the baculovirus with the DNA construct encoding the chimeric protein. Wild type baculovirus which do not contain foreign DNA express the polyhedrin protein in the nuclei of the infected insect cells. However, the recombinants do not produce polyhedrin protein and do not produce viral occlusions. Instead, the recombinants produce the chimeric protein.

[0156] Alternative insect host cells for use with this expression system are Sf21 cell line derived from Spodoptera frugiperda and High Five cell lines derived from Trichoplusia ni.

[0157] Other acceptable cloning vectors include phages, cosmids or artificial chromosomes. For example, bacteriophage lambda is a useful cloning vector. This phage can accept pieces of foreign DNA up to about 20,000 base pairs
in length. The lambda phage genome is a linear double stranded DNA molecule with single stranded complementary
(cohesive) ends which can hybridize with each other when inside an infected host cell. The lambda DNA is cut with a
restriction endonuclease and the foreign DNA, e.g. the DNA to be cloned, is ligated to the phage DNA fragments. The
resulting recombinant molecule is then packaged into infective phage particles. Host cells are infected with the phage

particles containing the recombinant DNA. The phage DNA replicates in the host cell to produce many copies of the desired DNA sequence.

[0158] Cosmids are hybrid plasmid/bacteriophage vectors which can be used to clone DNA fragments of about 40,000 base pairs. Cosmids are plasmids which have one or more DNA sequences called "cos" sites derived from bacteriophage lambda for packaging lambda DNA into infective phage particles. Two cosmids are ligated to the DNA to be cloned. The resulting molecule is packaged into infective lambda phage particles and transfected into bacteria host cells. When the cosmids are inside the host cell they behave like plasmids and multiply under the control of a plasmid origin of replication. The origin of replication is a sequence of DNA which allows a plasmid to multiply within a host cell.

**[0159]** Yeast artificial chromosome vectors are similar to plasmids but allow for the incorporation of much larger DNA sequences of about 400,000 base pairs. The yeast artificial chromosomes contain sequences for replication in yeast. The yeast artificial chromosome containing the DNA to be cloned is transformed into yeast cells where it replicates thereby producing many copies of the desired DNA sequence. Where phage, cosmids, or yeast artificial chromosomes are employed as cloning vectors, expression of the chimeric protein may be obtained by culturing host cells that have been transfected or transformed with the cloning vector in a suitable culture medium.

10

20

30.

35

40

45

55

[0160] Chimeric proteins disclosed herein are intended for use in treating mammals or other animals. The therapeutically active moieties described above, e.g., osteogenic agents such as BMPs, TGFs, decorin, and/or fragments of each of them, are all to be considered as being or having been derived from physiologically active agents for purposes of this description. The chimeric proteins and DNA constructs which incorporate a domain derived from one or more cellular physiologically active agents can be used for in vivo therapeutic treatment, in vitro research or for diagnostic purposes in general.

[0161] When used in vivo, formulations containing the present chimeric proteins may be placed in direct contact with viable tissue, including bone, to induce or enhance growth, repair and/or replacement of such tissue. This may be accomplished by applying a chimeric protein directly to a target site during surgery. It is contemplated that minimally invasive techniques such as endoscopy are to be used to apply a chimeric protein to a desired location. Formulations containing the chimeric proteins disclosed herein may consist solely of one or more chimeric proteins or may also incorporate one or more pharmaceutically acceptable adjuvants.

[0162] In an alternate embodiment, any of the above-described chimeric proteins may be contacted with, adhered to, or otherwise incorporated into an implant such as a drug delivery device or a prosthetic device. Chimeric proteins may be microencapsulated or macroencapsulated by liposomes or other membrane forming materials such as alginic acid derivatives prior to implantation and then implanted in the form of a pouchlike implant. The chimeric protein may be microencapsulated in structures in the form of spheres, aggregates of core material embedded in a continuum of wall material or capillary designs. Microencapsulation techniques are well known in the art and are described in the Encyclopedia of Polymer Science and Engineering, Vol. 9, pp. 724 et seq. (1980) hereby incorporated herein by reference.

[0163] Chimeric proteins may also be coated on or incorporated into medically useful materials such as meshes, pads, felts, dressings or prosthetic devices such as rods, pins, bone plates, artificial joints, artificial limbs or bone augmentation implants. The implants may, in part, be made of biocompatible materials such as glass, metal, ceramic, calcium phosphate or calcium carbonate based materials. Implants having biocompatible biomaterials are well known in the art and are all suitable for use herein. Implant biomaterials derived from natural sources such as protein fibers, polysaccharides, and treated naturally derived tissues are described in the Encyclopedia of Polymer Science and Engineering, Vol. 2, pp. 267 et seq. (1989) hereby incorporated herein by reference. Synthetic biocompatible polymers are well known in the art and are also suitable implant materials. Examples of suitable synthetic polymers include urethanes, olefins, terephthalates, acrylates, polyesters and the like. Other acceptable implant materials are biodegradable hydrogels or aggregations of closely packed particles such as polymethylmethacrylate beads with a polymerized hydroxyethyl methacrylate coating. See the Encyclopedia of Polymer Science and Engineering, Vol. 2, pp. 267 et seq. (1989) hereby incorporated herein by reference.

[0164] The chimeric protein herein provides a useful way for immobilizing or coating a physiologically active agent on a pharmaceutically acceptable vehicle to deliver the physiologically active agent to desired sites in viable tissue. Suitable vehicles include those made of bioabsorbable polymers, biocompatible nonabsorbable polymers, lactoner putty and plaster of Paris. Examples of suitable bioabsorbable and biocompatible polymers include homopolymers, copolymers and blends of hydroxyacids such as lactide and glycolide, other absorbable polymers which may be used alone or in combination with hydroxyacids including dioxanones, carbonates such as trimethylene carbonate, lactones such as caprolactone, polyoxyalkylenes, and oxylates. See the Encyclopedia of Polymer Science and Engineering, Vol. 2, pp. 230 et seq. (1989) hereby incorporated herein by reference.

[0165] These vehicles may be in the form of beads, particles, putty, coatings or film vehicles. Diffusional systems in which a core of chimeric protein is surrounded by a porous membrane layer are other acceptable vehicles.

[0166] In another aspect, the amount of amino acid analog(s) transport into a target cell can be regulated by con-

trolling the tonicity of the growth media. A hypertonic growth media increases uptake of *trans*-4-hydroxyproline into *E. coli* as illustrated in Figure 2A. All known methods of increasing osmolality of growth media are appropriate for use herein including addition of salts such as sodium chloride, KCl, MgCl<sub>2</sub> and the like, and sugars such as sucrose, glucose, maltose, etc. and polymers such as polyethylene glycol (PEG), dextran, cellulose, etc. and amino acids such as glycine. Increasing the osmolality of growth media results in greater intracellular concentration of amino acid analog (s) and a higher degree of complexation of amino acid analog(s) to tRNA. As a consequence, proteins produced by the cell achieve a higher degree of incorporation of amino acid analogs. Figure 12 illustrates percentage of incorporation of proline and hydroxyproline into MBP under isotonic and hypertonic media conditions in comparison to proline in native MBP. Thus, manipulating osmolality, in addition to adjusting concentration of amino acid analog(s) in growth media allows a dual-faceted approach to regulating their uptake into prokaryotic cells and eukaryotic cells as described above and consequent incorporation into target polypeptides.

[0167] Any growth media can be used herein including commercially available growth media such as M9 minimal medium (available from Gibco Life Technologies, Inc.), LB medium, NZCYM medium, terrific broth, SOB medium and others that are well known in the art.

15

20

25

35

40

45

50

55

[0168] Collagen from different tissues can contain different amounts of trans-4-hydroxyproline. For example, tissues that require greater strength such as bone contain a higher number of trans-4-hydroxyproline residues than collagen in tissues requiring less strength, e.g., skin. The present system provides a method of adjusting the amount of trans-4-hydroxyproline in collagen, collagen fragments, collagen-like peptides, and chimeric peptides having a collagen domain, collagen fragment domain or collagen-like peptide domain fused to a physiologically active domain, since by increasing or decreasing the concentration of trans-4-hydroxyproline in growth media, the amount of trans-4-hydroxyproline incorporated into such polypeptides is increased or decreased accordingly. The collagen, collagen fragments, collagen-like peptides and above-chimeric peptides can be expressed with predetermined levels of trans-4-hydroxyproline. In this manner physical characteristics of an extracellular matrix can be adjusted based upon requirements of end use. Without wishing to be bound by any particular theory, it is believed that incorporation of trans-4-hydroxyproline into the EMP moieties herein provides a basis for self aggregation as described herein.

**[0169]** In another aspect, the combination of incorporation of *trans*-4-hydroxyproline into collagen and fragments thereof using hyperosmotic media and genes which have been altered such that codon usage more closely reflects that found in *E. coli*, but retaining the amino acid sequence found in native human collagen, surprisingly resulted in production by *E. coli* of human collagen and fragments thereof which were capable of self aggregation.

[0170] The human collagen Type I ( $\alpha_1$ ) gene sequence (Figure 27A-27E) (SEQ. ID. NO. 15) contains a large number of glycine and proline codons (347 glycine and 240 proline codons) arranged in a highly repetitive manner. Table I below is a codon frequency tabulation for the human Type I ( $\alpha_1$ ) collagen gene. Of particular note is that the GGA glycine codon occurs 64 times and the CCC codon for proline occurs 93 times. Both of these codons are considered to be rare codons in *E. coli.* See, Sharp, P.M. and W.-H. Li. Nucleic Acids Res. 14: 7737-7749, 1986. These, and similar considerations for other human collagen genes are shown herein to account for the difficulty in expressing human collagen genes in *E. coli.* 

TABLE 1

							'					
	Codon	Count	%age	Codon	Count	%age	Codon	Count	%age	Codon	Count	%age
	TTT-	1	0.09	ТСТ-	18	1.70	TAT-	2	0.18	TGT-	0	0.00
	Phe			Ser			Tyr			Cys		
	TTC-	14	1.32	тсс-	4	0.37	TAC-	2	0.18	TGC-	0	0.00
	Phe			Ser			Tyr			Cys		
	TTA-	0	0.00	TCA-	2	0.18	TAA-	0	0.00	TGA-***	0	0.00
	Leu			Ser			***					
	TTG-	3	0.28	TCG-	0	0.00	TAG-	0	0.00	TGG-	0	0.00
	Leu			Ser			***			Trp		_
	CTT-	4	0.37	ССТ-	141	13.33	CAT-	0	0.00	CGT-	26	2.45
!	Leu			Pro			His			Arg		
	CTC-	7	0.66	CCC-	93	8.79	CAC-	3	0.28	CGC-	6	0.56
	Leu			Pro			His			Arg		
	CTA-	0	0.00	CCA-	6	.0.56	CAA-	13	1.22	CGA-	11	1.04
;	Leu			Pro			Gin			Arg		
	CTG-	7	0.66	CCG-	0	0.00	CAG-	17	1.60	CGG-	1	0.09
	Leu			Pro			Gln			Arg		

TABLE 1 (continued)

5

∂. 10

15

20

25

30

45

50

55

Codon	Count	%age									
ATT- lle	6	0.56	ACT- Thr	11	1.04	AAT- Asn	6	0.56	AGT- Ser	4	0.37
ATC- lle	0	0.00	ACC- Thr	4	0.37	AAC- Asn	5	0.47	AGC- Ser	11	1.04
ATA- Ile	1	0.09	ACA- Thr	2	0.18	AAA- Lys	19	1.79	AGA- Arg	9	0.85
ATG- Met	7	0.66	ACG- Thr	0	0.00	AAG- Lys	19	1.79	AGG- Arg	0	0.00
GTT- Val	10	0.94	GCT- Ala	93	8.79	GAT- Asp	23	2.17	GGT- Gly	174	16.46
GTC- Val	5	0.47	GCC- Ala	24	2.27	GAC- Asp	11	1.04	GGC- Gly	97	9.17
GTA- Val	0	0.00	GCA- Ala	6	0.56	GAA- Glu	24	2.27	GGA- Gly	64	6.05
GTG- Val	5	0.47	GCG- Ala	0	0.00	GAG- Glu	25	2.36	GGG- Gly	11	1.04

[0171] In a first step, the sequence of the heterologous collagen gene is changed to reflect the codon bias in E. coli as given in codon usage tables (e.g. Ausubel et al., (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York, New York; Wada et al., 1992, supra). Rare E. coli codons (See, Sharp, P.M. and W.-H. Li. Nucleic Acids Res. 14: 7737-7749, 1986) are avoided. Second, unique restriction enzyme sites are chosen that are located approximately every 120-150 base pairs in the sequence. In certain cases this entails altering the nucleotide sequence but does not change the amino acid sequence. Third, oligos of approximately 80 nucleotides are synthesized such that when two such oligos are annealed together and extended with a DNA polymerase they reconstruct a approximately 120-150 base pair section of the gene (Figure 28). The section of the gene encoding the very amino terminal portion of the protein has an initiating methionine (ATG) codon at the 5' end and a unique restriction site followed by a stop (TAAT) signal at the 3' end. The remaining sections have unique restriction sites at the 5' end and unique restriction sites followed by a TAAT stop signal the 3' end. The gene is assembled by sequential addition of each section to the preceding 5' section. In this manner, each successively larger section can be independently constructed and expressed. Figure 28 is a schematic representation of the construction of the human collagen gene starting from synthetic oligos. [0172] A fragment of the human Type I all collagen chain fused to the C-terminus of glutathione S-transferase (GST-D4, Fig. 29) (SEQ. ID. NO. 18) was prepared and tested for expression in E. coli strain JM109 (F-) under conditions of hyperosmotic shock. The collagen fragment included the C-terminal 193 amino acids of the triple helical region and the 26 amino acid C-terminal telopeptide. Fig. 29 is a schematic of the amino acid sequence of the GST-CoIECoI (SEQ. ID. NO. 17) and GST-D4 (SEQ. ID. NO. 18) fusion proteins. ColECol comprises the 17 amino acid N-terminal telopeptide, 338 Gly-X-Y repeating tripeptides, and the 26 amino acid C-terminal telopeptide. There is a unique methionine at the junction of GST and D4, followed by 64 Gly-X-Y repeats, and the 26 amino acid telopeptide. The residue (Phel99) in the C-terminal telopeptide of D4 where pepsin cleaves is indicated. The gene was synthesized for the collagen fragment from synthetic oligonucleotides designed to reflect optimal E. coli usage. Fig. 30 is a table depicting occurrence of the four proline and four glycine codons in the human Type I α1 gene (HCoI) and the Type I α1 gene with optimized E. coli codon usage (CoIECol). Usage of the remaining codons in CoIECol was also optimized for E. coli expression according to Wada et al., supra. Protein GST-D4 was efficiently expressed in JM109 (F-) in minimal media lacking proline but supplemented with Hyp and Nacl (See Figs. 31 and 32). Expression was dependent on induction with isopropyl-1-thio-β-galactopyranoside (IPTG), trans-4-hydroxyproline and NaCl. At a fixed Nacl concentration of 500 mM, expression was minimal at trans-4-hydroxyproline concentrations below ~20 mM while the expression level plateaued at trans-4-hydroxyproline concentrations above 40 mM. See Fig. 31 which depicts a gel showing expression and dependence of expression of GST-D4 on hydroxyproline. The concentration of hydroxyproline is indicated above each lane. Osmolyte (NaCl) was added at 500 mM in each culture and each was induced with 1.5 mM IPTG. The arrow marks the position of GST-D4. Likewise, at a fixed trans-4-hydroxyproline concentration of 40 mM, NaCl concentrations below 300 mM resulted in little protein accumulation and expression decreased above 700-800 mM NaCl. See Fig. 32 which depicts a gel showing expression of GST-D4 in hyperosmotic media. Lanes 2 and 3 are uninduced and induced samples, respectively, each without added osmolyte. The identity and quantity of osmolyte is indicated above each of the other lanes. Trans-4-Hydroxyproline was added at 40mM in each culture and all cultures except that in lane 1 were induced with 1.5 mM IPTG. The arrow marks the position of GST-D4.

5

10

15

20

30

35

45

55

[0173] Either sucrose or KC1 can be substituted for NaC1 as the osmolyte (See Fig. 32). Thus, the osmotic shock-mediated intracellular accumulation of *trans*-4-hydroxyproline was a critical determinant of expression rather than the precise chemical identity of the osmolyte. Despite the large number of prolines (66) in GST-D4, its size (46 kDA), and non-optimal growth conditions, it was expressed at ~10% of the total cellular protein. Expressed proteins of less than full-length indicative of aborted transcription, translation, or mRNA instability were not detected.

[0174] The gene for protein D4 contains 52 proline codons. In the expression experiments reflected in Figs. 31 and 32, it was expected that trans-4-hydroxyproline would be inserted at each of these codons resulting in a protein where trans-4-hydroxyproline had been substituted for all prolines. To confirm this, GST-D4 was cleaved with BrCN in 0.1 N HC1 at methionines within GST and at the unique methionine at the N-terminal end of D4, and D4 purified by reverse phase HPLC. Crude GST-D4 was dissolved in 0.1 M HC1 in a round bottom flask with stirring. Following addition of a 2-10 fold molar excess of clear, crystalline BrCN, the flask was evacuated and filled with nitrogen. Cleavage was allowed to proceed for 24 hours, at which time the solvent was removed in vacuo. The residue was dissolved in 0.1% trifluoroacetic acid (TFA) and purified by reverse-phase HPLC using a Vydac C4 RP-HPLC column (10 x 250 mm, 5 μ, 300 Å) on a BioCad Sprint system (Perceptive Biosystems, Framingham, MA). D4 was eluted with a gradient of 15 to 40% acetonitrile/0.1% TFA over a 45 min. period. D4 eluted as a single peak at 26% acetonitrile/0.1% TFA. Standard BrCN cleavage conditions (70% formic acid) resulted in extensive formylation of D4, presumably at the hydroxyl groups of the trans-4-hydroxyproline residues. Formylation of BrCN/formic acid-cleaved proteins had been noted before (Beavis et al., Anal. Chem., 62, 1836 (1990)). Amino acid analysis was carried out on a Beckman ion exchange instrument with post-column derivatization. N-terminal sequencing was performed on an Applied Biosystems sequencer equipped with an on-line HLPC system. Electrospray mass spectra were obtained with a VG Biotech BIO-Q quadropole analyzer by M-Scan, Inc. (West Chester, PA). For CD thermal melts, the temperature was raised in 0.5°C increments from 4°C to 85°C with a four minute equilibration between steps. Data were recorded at 221.5 nm. The thermal transition was calculated using the program ThermoDyne (MORE). The electrospray mass spectroscopy of this protein gave a single molecular ion corresponding to a mass of 20,807 Da. This mass is within 0.05% of that expected for D4 if it contains 100% trans-4-hydroxyproline in lieu of proline. Proline was not detected in amino acid analysis of purified D4, again consistent with complete substitution of trans-4-hydroxyproline for proline. To confirm further that trans-4-hydroxyproline substitution had only occurred at proline codons, the N-terminal 13 amino acids of D4 was sequenced as above. The first 13 codons of D4 specify the protein sequence H<sub>2</sub>N-Gly-Pro-Pro-Gly-Leu-Ala-Gly-Pro-Pro-Gly-Glu-Ser-Gly (SEQ. ID. NO. 41). The sequence found was H<sub>2</sub>N-Gly-Hyp-Hyp-Gly-Leu-Ala-Gly-Hyp-Hyp-Gly-Glu-Ser-Gly (SEQ. ID. NO. 42), see Fig. 69. Taken together, these results indicate that trans-4-hydroxyproline (Hyp) was inserted only at proline codons and that the fidelity of the E. coli translational machinery was not otherwise altered by either the high intracellular concentration or trans-4-hydroxyproline or hyperosmotic culture conditions.

[0175] To determine whether D4, containing trans-4-hydroxyproline in both the X and Y positions, forms homotrimeric helices and to compare stability to native collagen, the following was noted: In neutral pH phosphate buffer, D4 exhibits a circular dichroism (CD) spectrum characteristic of a triple helix (See Fig. 33 and Bhatnagar et al., Circular Dichroism and the Conformational Analysis of Biomolecules, G.D. Fasman, Ed. Plenum Press, New York, (1996 p. 183). Fig. 33 illustrates circular dichroism spectra of native and heat-denatured D4 in neutral phosphate buffer. HPLC-purified D4 was dissolved in 0.1M sodium phosphate, pH 7.0, to a final concentration of 1 mg/mL (E<sup>280</sup>=3628 M<sup>-1</sup>·cm<sup>-1</sup>). The solution was incubated at 4°C for two days to allow triple helices to form prior to analysis. Spectra were obtained on an Aviv model 62DS spectropolarimeter (Yale University, Molecular Biophysics and Biochemistry Department), A 1 mm path length quartz suprasil fluorimeter cell was used. Following a 10 min. incubation period at 4°C, standard wavelength spectra were recorded from 260 to 190 nm using 10 sec acquisition times and 0.5 nm scan steps. This spectrum is characterized by a negative ellipticity at 198 nm and a positive ellipticity at 221 nm. The magnitudes of both of these absorbances was greater in neutral pH buffer compared to acidic conditions. Comparable dependence of stability on pH has been noted for collagen-like triple helices. See, e.g., Venugopal et al., Biochemistry, 33, 7948 (1994). Heating at 85°C for five minutes prior to obtaining the CD spectrum decreased the magnitude of the absorbance at 198 nm and abolished the absorbance at 221 nm (Fig. 33). This behavior is also typical of the triple helical structure of collagen. See, R.S. Bhatnagar et al., Circular Dichroism and the Conformational Analysis of Biomolecules G.D. Fasman, Ed., supra. A thermal melt profile of D4 conducted as above in phosphate buffer gave a melting temperature of about 29°C. A fragment of the C-terminal region of the bovine Type I α1 collagen chain comparable in length to D4 forms homotrimeric helices with a melting temperature of 26°C. (See, A. Rossi, et al., Biochemistry 35, 6048 (1996)). [0176] Resistance to pepsin digestion is a second commonly used indication of triple helical structure. At 4°C, the majority of D4 is digested rapidly by pepsin to a protein of slightly lower molecular weight. Fig. 34 is a gel illustrating the result of digestion of D4 with bovine pepsin. Purified D4 was dissolved in 0.1 M sodium phosphate, pH 7.0, to 1.6 μg/μl and incubated at 4°C for 7 days. Aliquots (10 μl) were placed into 1.5 ml centrifuge tubes and adjusted with water and 1 M acetic acid solutions to 25 µl final volume and 200 mM final acetic acid concentration. Each tube was then incubated for 20 min. at the indicated temperature and pepsin (0.5 μl of a 0.25 μg/μl solution) was added to each tube

and digestion allowed to proceed for 45 minutes. Following digestion, samples were quenched with loading buffer and analyzed by SDS-PAGE. However, the initial pepsin cleavage product is resistant to further digestion up to ~30°C. Amino terminal sequencing as above of the initial pepsin cleavage product showed that the N-terminus was identical to that of full-length D4. Mass spectral analysis as above of the digestion product gave a parent ion with a molecular weight consistent with cleavage in the C-terminal telopeptide on the N-terminal side of Phe119 (See Fig. 29) suggesting that this portion of the protein is either globular or of ill-defined structure and rapidly cleaved by pepsin while the triple helical region is resistant to digestion. Thus, despite global *trans*-4-hydroxyproline for proline substitution in both the X and Y positions, D4 formed triple helices of stability similar to comparably sized fragments of bovine collagen containing Hyp at the normal percentage and only in the Y position.

10

15

20

25

30

35

40

50

[0177] The full-length human Type I α1 collagen chain, although more than four times the size of D4, also expressed as a N-terminal fusion with GST (GST-ColECol, Fig. 29) in JM109(F-) in Hyp/NaCl media. Fig. 35 is a gel depicting expression of GST-HCol and GST-ColECol. *Trans*-4-hydroxyproline was added at 40 mM and NaCl at 500 mM. Expression was induced with 1.5 mM IPTG. The arrow marks the position of GST-ColECol. In the procedures resulting in the gels shown in Figs. 31, 32 and 35, five ml cultures of JM109 (F-) harboring the expression plasmid in LB media containing 100 μg/ml ampicillin were grown overnight. Cultures were centrifuged and the cell pellets washed twice with five ml of M9/Amp media (See, J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989)) supplemented with 0.5% glucose and 100 μg/ml of all amino acids except glycine and alanine which were at 200 μg/ml and containing no proline. The cells were finally resuspended in five ml of the above media. Following incubation at 37°C for 30 min., hydroxyproline, osmolyte, or IPTG were added as indicated. After four hours, aliquots of the cultures were analyzed by SDS-PAGE.

[0178] Like D4, the gene for protein CoIECoI was constructed from synthetic oligonucleotides designed to mimic codon usage in highly-expressed E. coli genes. In contrast to GST-ColECol, expression from a GST-human Type I α1 gene fusion (pHCol) identical to GST-ColECol in coded amino acid sequence but containing the human codon distribution could not be detected in Coomassie blue-stained SDS-PAGE gels of total cell lysates of induced JM109 (F-)/ pHCol cultures (Fig. 35). The gene for the Type I α1 collagen polypeptide was cloned by polymerase chain reaction of the gene from mRNA isolated from human foreskin cells (HS27, ATCC 1634) with primers designed from the published gene sequence (GenBank Z74615). The 5' primer added a flanking EcoR I recognition site and the 3' primer a flanking Hind III recognition site. The gene was cloned into the EcoR I/Hind III site of plasmid pBSKS+ (Stratagene, La Jolla, CA), four mutations corrected using the ExSite mutagenesis kit (Stratagene, La Jolla, CA), the sequence confirmed by dideoxy sequencing, and finally the EcoR I/Xho I fragment subcloned into plasmid pGEX-4T.1 (Pharmacia, Piscataway, NJ). The GST-HCol gene is expression-competent because a protein of the same molecular weight as GST-ColECol is detected when immunoblots of total cell lysates are probed with an anti-Type I collagen antibody. Thus, sequence or structural differences between the genes for ColECol and HCol are critical determinants of expression efficiency in E. coli. This is likely due to the codon distribution in these genes and ultimately to differences in tRNA isoacceptor levels in E. coli compared to humans. GST-ColECol, GST-D4, and GST-HCol do not accumulate in hyperosmotic shock media when proline is substituted for hydroxyproline or in rich media. A possible explanation is that the trans-4-hydroxyproline-containing proteins may be resistant to degradation because they fold into a protease-resistant triple helix while the proline-containing proteins do not adopt this structure. The large number of codons non-optimal for E. coli found in the human gene and the instability of proline-containing collagen in E. coli may, in part, explain why expression of human collagen in E. coli has not been previously reported.

[0179] As discussed above, collagen mimetic polypeptides, i.e., engineered polypeptides having certain compositional and structural traits in common with collagen are also provided herein. Such collagen mimetic polypeptides may also be made to incorporate amino acid analogs as described above. GST-CM4 consists of glutathione S-transferase fused to 30 repeats of a Gly-X-Y sequence. The Gly-X-Y repeating section mimics the Gly-X-Y repeating unit of human collagen and is referred to as collagen mimetic 4 or CM4 herein. Thus, the hydroxyproline-incorporating technology was also demonstrated to work with a protein and DNA sequence analogous to that found in human collagen. Amino acid analysis of purified CM4 protein express in *E. coli* strain JM109 (F-) under hydroxyproline-incorporating conditions compared to analysis of the same protein expressed under proline-incorporating conditions, demonstrates that the techniques herein result in essentially complete substitution of hydroxyproline for proline. The amino acid analysis was performed on CM4 protein that had been cleaved from and purified away from GST. This removes any possible ambiguities associated with the fusion protein.

[0180] Expression in media containing at least about 200 mM NaCl is preferable to accumulate significant amount of protein containing hydroxyproline. A concentration of about 400-500 mM NaCl appears to be optimal. Either KCl, sucrose or combinations thereof may be used in substitution of or with NaCl. However, expression in media without an added osmolyte (i.e. under conditions that more closely mimic those of Deming et al., In Vivo Incorporation of Proline Analogs into Artificial Protein, Poly. Mater. Sci. Engin. Proceed., supra.) did not result in significant expression of hydroxyproline-containing proteins in JM109 (F-). This is illustrated in Figure 36 which is a scan of a SDS-PAGE get showing the expression of GST-CM4 in media with or without 500 mM Nacl and containing either proline or hydroxy-

5

10

15

20

25

35

45

55

proline. The SDS-PAGE gel reflects 5 hour post-induction samples of GST-CM4 expressed in JM109 (F-). Equivalent amounts, based on OD600nm, of each culture were loaded in each lane. Gels were stained with Coomasie Blue, destained, and scanned on a PDI 420oe scanner. Lane 1: 2.5mM proline/0mM NaCl. Lane 2: 2.5mM proline/500mM NaCl. Lane 3: 80mM hydroxyproline/0mM NaCl. Lane 4: 80mM hydroxyproline/500mM NaCl. Lane 5: Molecular weight markers. The lower arrow indicates the migration position of proline-containing GST-CM4 in lanes 1 and 2. The upper arrow indicates the migration position of hydroxyproline-containing GST-CM4 in lanes 3 and 4. Note that GST-CM4 expressed in the presence of hydroxyproline runs at a higher apparent molecular weight (compare lanes 1 and 4). This is expected since hydroxyproline is of greater molecular weight than proline. If all the prolines in GST-CM4 are substituted with hydroxyproline, the increase in molecular weight is 671 Da (+2%). Note also that protein expressed in the presence of proline accumulates in cultures irrespective of the NaCl concentration (compare lanes 1 and 2). In contrast, significant expression in the presence of hydroxyproline only occurs in the culture containing 500 mM NaCl (compare lanes 3 and 4). Figure 37 further illustrates the dependence of expression on Nacl concentration by showing that significant expression of GST-CM4 occurs only at Nacl concentration greater than 200 mM. The SDS-PAGE gel reflects 6 hour post-induction samples of GST-CM4 expressed in JM109 (F-) with varying concentrations of NaCl. All cultures contained 80 mM hydroxyproline. Lane 1: 500 mM NaCl, not induced. Lanes 2-6: 500 mM, 400 mM, 300 mM, 200 mM, and 100 mM NaCI, respectively. All induced with 1.5 mM IPTG. Lane 7: Molecular weight markers. The arrow indicates the migration position of hydroxyproline-containing GST-CM4. Figure 38 is a scan of an SDS-PAGE gel of expression of GST-CM4 in either 400 mM NaCl or 800 mM sucrose. The SDS-PAGE gel reflects 4 hour post-induction samples of GST-CM4 expressed in JM109 (F-). All cultures contained 80 mM hydroxyproline and all, except that electrophoresed in lane 2, contained 400 mM NaCl. Lane 2 demonstrates expression in sucrose in lieu of NaCl. Lane 1: Molecular weight markers. Lane 2: 800 mM sucrose (no NaCl). Lanes 3-9: 0 mM, 0.025 mM, 0.1 mM, 0.4 mM, 0.8 mM, 1.25 mM, 2.5 mM proline, respectively. The upper arrow indicates the migration position of hydroxyproline-containing GST-CM4 and the lower arrow indicates the migration position of proline-containing GST-CM4. Expression is apparent in both cases (compare lanes 2 and 3).

[0181] If expression of GST-CM4, as described in Example 17 below, is performed in varying ratios of hydroxyproline and proline the expressed protein appears to contain varying amounts of hydroxyproline. Thus, if only hydroxyproline is present during expression, a single expressed protein of the expected molecular weight is evident on a SDS-PAGE gel (Figure 38, lane 3). If greater than approximately 1 mM proline is present, again a single expressed protein is evident, but at a lower apparent molecular weight, as expected for the protein containing only proline (Figure 38, lanes 7-9). If lesser amount of proline are used during expression, species of apparent molecular weight intermediate between these extremes are evident. This phenomenon, evident as a "smear" or "ladder" of proteins running between the two molecular weight extremes on an SDS-PAGE gel, is illustrated in lanes 3-9 of Figure 38. Lanes 3-9 on this gel are proteins from expression in a fixed concentration of 80 mM hydroxyproline and 400 mM NaCl. However, in moving from lane 3 to 9 the proline concentration increases from none (lane 3) to 2.5 mM (lane 9) and expression shifts from a protein of higher molecular weight (hydroxyproline-containing GST-CM4) to lower molecular weight (proline-containing GST-CM4). At proline concentrations of 0.025 mM and 0.1 mM, species of intermediate molecular weight are apparent (lanes 4 and 5). This clearly demonstrates that the percent incorporation of hydroxyproline in an expressed protein can be controlled by expression in varying ratios of analogue to amino acid.

[0182] Proline starvation prior to hydroxyproline incorporation is an important technique used herein. It insures that no residual proline is present during expression to compete with hydroxyproline. This enables essentially 100% substitution with the analogue. As shown in Figure 38, starvation conditions allow expression under precisely controlled ratios of proline and hydroxyproline. The amount of hydroxyproline vs. proline incorporated into the recombinant protein can therefore be controlled. Thus, particular properties of the recombinant protein that depend upon the relative amount of analogue incorporated can be tailored by the present methodology to produce polypeptides with unique and beneficial properties.

[0183] Human collagen, collagen fragments, collagen-like peptides (collagen mimetics) and the above chimeric polypeptides produced by recombinant processes have distinct advantages over collagen and its derivatives obtained from non-human animals. Since the human gene is used, the collagen will not act as a xenograft in the context of a medical implant. Moreover, unlike naturally occurring collagen, the extent of proline hydroxylation can be predetermined. This unprecedented degree of control permits detailed investigation of the contribution of *trans*-4-hydroxyproline to triple helix stabilization, fibril formation and biological activity. In addition, design of medical implants based upon the desired strength of collagen fibrils is enabled.

[0184] The following examples are included for purposes of illustration and are not to be construed as limitations herein.

### **EXAMPLE 1**

Trans-membrane Transport

[0185] A 5 mL culture of E. coli strain DH5α (supE44 ΔlacU169 (φ80lacZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) containing a plasmid conferring resistance to ampicillin (pMAL-c2, Fig. 1) was grown in Luria Broth to confluency (~16 hours from inoculation). These cells were used to inoculate a 1 L shaker flask containing 500 mL of M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100 μg/mL ampicillin supplemented with all amino acids at 20 μg/mL) which was grown to an AU<sub>600</sub> of 1.0 (18-20 hours). The culture was divided in half and the cells harvested by centrifugation. The cells from one culture, were resuspended in 250 mL M9 media and those from the other in 250 mL of M9 media containing 0.5M NaCl. The cultures were equilibrated in an air shaker for 20 minutes at 37 °C (225 rpm) and divided into ten 25 mL aliquots. The cultures were returned to the shaker and 125 μl of 1M hydroxyproline in distilled H<sub>2</sub>0 was added to each tube. At 2, 4, 8, 12, and 20 minutes, 4 culture tubes (2 isotonic, 2 hypertonic) were vacuum filtered onto 1 μm polycarbonate filters that were immediately placed into 2 mL microfuge tubes containing 1.2 mL of 0.2M NaOH/2% SDS in distilled H<sub>2</sub>0. After overnight lysis, the filters were carefully removed from the tubes, and the supernatant buffer was assayed for hydroxyproline according to the method of Grant, Journal of Clinical Pathology, 17:685 (1964). The intracellular concentration of trans-4-hydroxyproline versus time is illustrated graphically in Figure 2.

# 20 EXAMPLE 2

25

30

35

45

50

Effects of Salt Concentration on Transmembrane Transport

[0186] To determine the effects of salt concentration on transmembrane transport, an approach similar to Example 1 was taken. A 5 mL culture of *coli* strain DH5 $\alpha$  (supE44  $\Delta lac$ U169 ( $\phi 80 lacZ \Delta M15$ ) hsdR17 recA1 ental gyrA96 thi-1 relA1) containing a plasmid conferring resistance to ampicillin (pMAL-c2, Fig. 1) was grown in Luria Broth to confluency (~16 hours from inoculation). These cells were used to inoculate a 1 L shaker flask containing 500 mL of M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100  $\mu$ g/mL ampicillin supplemented with all amino acids at 20  $\mu$ g/mL) that was then grown to an AU $_{600}$  of 0.6. The culture was divided into three equal parts, the cells in each collected by centrifugation and resuspended in 150 mL M9 media, 150 mL M9 media containing 0.5M NaCl, and 150 mL M9 media containing 1.0M NaCl, respectively. The cultures were equilibrated for 20 minutes on a shaker at 37° C (225rpm) and then divided into six 25 mL aliquots. The cultures were returned to the shaker and 125  $\mu$ L of 1M hydroxyproline in distilled H<sub>2</sub>0 was added to each tube. At 5 and 15 minutes, 9 culture tubes (3 isotonic, 3 x 0.5M NaCl, and 3 x 1.0M NaCl) were vacuum filtered onto 1  $\mu$ m polycarbonate filters that were immediately placed into 2 mL microfuge tubes containing 1.2 mL of 0.2M NaOH/2% SDS in distilled H<sub>2</sub>0. After overnight lysis, the filters were removed from the tubes and the supernatant buffer assayed for hydroxyproline according to the method of Grant, supra.

# **EXAMPLE 2A**

40 Effects of Salt Concentration on Transmembrane Transport

[0187] To determine the effects of salt concentration on transmembrane transport, an approach similar to Example 1 was taken. A saturated culture of JM109 (F-) harboring plasmid pD4 (Fig. 48) growing in Luria Broth (LB) containing 100μg/ml ampicillin (Amp) was used to inoculate 20 ml cultures of LB/Amp to an OD at 600 nm of 0.1 AU. The cultures were grown with shaking at 37°C to an OD 600 nm between 0.7 and 1.0 AU. Cells were collected by centrifugation and washed with 10 ml of M9 media. Each cell pellet was resuspended in 20 ml of M9/Amp media supplemented with 0.5% glucose and 100μg/ml of all of the amino acids except proline. Cultures were grown at 37°C for 30 min. to deplete endogenous proline. After out-growth, Nacl was added to the indicated concentration, Hyp was added to 40mM, and IPTG to 1.5mM. After 3 hours at 37°C, cells from three 5 ml aliquots of each culture were collected separately on polycarbonate filters and washed twice with five ml of M9 media containing 0.5% glucose and the appropriate concentration of NaCl. Cells were lysed in 1 ml of 70% ethanol by vortexing for 30 min. at room temperature. Cell lysis supernatants were taken to dryness, resuspended in 100μl of 2.5 N NaOH, and assayed for Hyp by the method of Neuman and Logan, R.E. Neuman and M.A. Logan, Journal of Biological Chemistry, 184:299 (1950). Total protein was determined with the BCA kit (Pierce, Rockford II) after cell lysis by three sonication/freeze-thaw cycles. The data are the mean ± standard error of three separate experiments. The intracellular concentration of *trans*-4-hydroxyproline versus NaCl concentration is illustrated graphically in Figure 2A.

### **EXAMPLE 3**

5

10

15

20

25

30

35

45

50

55

Determination Of Proline Starvation Conditions in E. Coli

[0188] Proline auxotrophic *E. coli* strain NM519 (*pro*-) including plasmid pMAL-c2 which confers ampicillin resistance was grown in M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100μg mL ampicillin supplemented with all amino acids at 20 μg/mL except proline which was supplemented at 12.5 mg/L) to a constant AU<sub>600</sub> of 0.53 AU (17 hours post-inoculation). Hydroxyproline was added to 0.08<u>M</u> and hydroxyproline-dependent growth was demonstrated by the increase in the OD<sub>600</sub> to 0.61 AU over a one hour period.

# **EXAMPLE 4**

Hydroxyproline Incorporation Into Protein in E. coli Under Proline Starvation Conditions

[0189] Plasmid pMAL-c2 (commercially available from New England Biolabs) containing DNA encoding for maltose-binding protein (MBP) was used to transform proline auxotrophic *E. coli* strain NM519 (*pro*-). Two 1 L cultures of transformed NM519 (*pro*-) in M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100 μg/mL ampicillin supplemented with all amino acids at 20 μg/mL except proline which was supplemented at 12.5 mg/L) were grown to an AU<sub>600</sub> Of 0.53 (~17 hours post-inoculation). The cells were harvested by centrifugation, the media in one culture was replaced with an equal volume of M9 media containing 0.08M hydroxyproline and the media in the second culture was replaced with an equal volume of M9 media containing 0.08M hydroxyproline and 0.5M NaCl. After a one hour equilibration, the cultures were induced with 1mM isopropyl-β-D-thiogalactopyranoside. After growing for an additional 3.25 hours, cells were harvested by centrifugation, resuspended in 10 mL of 10mM Tris-HCl (pH 8), 1mM EDTA, 100mM NaCl (TEN buffer), and lysed by freezing and sonication. MBP was purified by passing the lysates over 4 mL amylose resin spin columns, washing the columns with 10 mL of TEN buffer, followed by elution of bound MBP with 2 mL of TEN buffer containing 10mM maltose. Eluted samples were sealed in ampules under nitrogen with an equal volume of concentrated HCl (11.7M) and hydrolysed for 12 hours at 120 °C. After clarification with activated charcoal, hydroxyproline content in the samples was determined by HPLC and the method of Grant, *supra*. The percent incorporation of *trans*-4-hydroxyproline compared to proline into MBP is shown graphically in Figure 12.

# **EXAMPLE 5**

Hydroxyproline Incorporation Into Protein in S. cerevisiae via Integrating Vectors Under Proline Starvation Conditions

[0190] The procedure described in Example 4 above is performed in yeast using an integrating vector which disrupts the proline biosynthetic pathway. A gene encoding human Type  $1(\alpha_1)$  collagen is inserted into a unique shuttle vector behind the inducible GAL10 promoter. This promoter/gene cassette is flanked by a 5' and 3' terminal sequence derived from a S. cerevisiae proline synthetase gene. The plasmid is linearized by restriction digestion in both the 5' and 3' terminal regions and used to transform a proline-prototrophic S. cerevisiae strain. The transformation mixture is plated onto selectable media and transformants are selected. By homologous recombination and gene disruption, the construct simultaneously forms a stable integration and converts the S. cerevisiae strain into a proline auxotroph. A single transformant is selected and grown at 30 °C in YPD media to an  $OD_{600}$  of 2 AU. The culture is centrifuged and the cells resuspended in yeast dropout media supplemented with all amino acids except proline and grown to a constant  $OD_{600}$  indicating proline starvation conditions. 0.08M L-hydroxyproline and 2% (w/v) galactose is then added. Cultures are grown for an additional 6-48 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Hydroxyproline-containing human Type 1( $\alpha_1$ ) collagen is purified by ammonium sulfate fractionation and column chromatography.

# **EXAMPLE 6**

Hydroxyproline Incorporation Into Protein in *S. cerevisiae* via Non-Integrating Vectors Under Proline Starvation Conditions

[0191] The procedure described above in Example 4 is performed in a yeast proline auxotroph using a non-integrating vector. A gene encoding human Type 1 ( $\alpha_1$ ) collagen is inserted behind the inducible GAL10 promoter in the YEp24 shuttle vector that contains the selectable Ura\* marker. The resulting plasmid is transformed into proline auxotrophic S. cerevisiae by spheroplast transformation. The transformation mixture is plated on selectable media and transformants are selected. A single transformant is grown at 30 °C in YPD media to an  $OD_{600}$  of 2 AU. The culture is centrifuged

and the cells resuspended in yeast dropout media supplemented with all amino acids except proline and grown to a constant  $OD_{600}$  indicating proline starvation conditions. 0.08M L-hydroxyproline and 2% (w/v) galactose is then added. Cultures are grown for an additional 6-48 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Hydroxyproline-containing human Type 1 ( $\alpha_1$ ) collagen is purified by ammonium sulfate fractionation and column chromatography.

### **EXAMPLE 7**

5

10

15

20

Hydroxyproline Incorporation Into Protein in a Baculovirus Expression System

[0192] A gene encoding human Type  $1(\alpha_1)$  collagen is inserted into the pBacPAK8 baculovirus expression vector behind the AcMNPV polyhedron promoter. This construct is co-transfected into SF9 cells along with linearized AcMNPV DNA by standard calcium phosphate co-precipitation. Transfectants are cultured for 4 days at 27 °C in TNM-FH media supplemented with 10 % FBS. The media is harvested and recombinant virus particles are isolated by a plaque assay. Recombinant virus is used to infect 1 liter of SF9 cells growing in Grace's media minus proline supplemented with 10% FBS and 0.08 M hydroxyproline. After growth at 27 °C for 2-10 days, cells are harvested by centrifugation and lysed by mechanical disruption.

Hydroxyproline-containing human Type 1 ( $\alpha_1$ ) collagen is purified by ammonium sulfate fractionation and column chromatography.

# **EXAMPLE 8**

Hydroxyproline Incorporation Into Human Collagen Protein in Escherichia coli Under Proline Starvation Conditions

25 [0193] A plasmid (pHuCol, Fig. 4) encoding the gene sequence of human Type I (α<sub>1</sub>) collagen (Figures 3A and 3B) (SEQ. ID. NO. 1) placed behind the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible tac promotor and also encoding β-lactamase is transformed into Escherichia coli proline auxotrophic strain NM519 (pro-) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 µg/ml ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 5 ml of LB containing 100 μg/ml ampicillin. After 30 growth for 10-16 hours with shaking (225 rpm) at 37 °C, this culture is used to inoculate 1 L of M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100 μg/mL ampicillin, supplemented with all amino acids at 20 μg/mL except proline which is supplemented at 12.5 mg/L) in a 1.5 L shaker flask. After growth at 37 °C, 225 rpm, for 15-20 hours post-inoculation, the optical density at 600 nm is constant at approximately 0.5 OD/mL. The cells are harvested by centrifugation (5000 rpm, 5 minutes), the media decanted, and the cells resuspended in 1 L of M9 minimal media containing 100 µg/mL ampicillin, 0.08M L-hydroxyproline, and 0.5M NaCl. Following growth for 1 hour at 37 °C, 225 rpm, IPTG is added to 1mM and the cultures allowed to grow for an additional 5-15 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Hydroxyproline-containing collagen is purified by ammonium sulfate fractionation and column chromatography.

# 40 EXAMPLE 9

Hydroxyproline Incorporation Into Fragments of Human Collagen Protein in Escherichia coli Under Proline Starvation Conditions

45 [0194] A plasmid (pHuCol-FI, Figure 6) encoding the gene sequence of the first 80 amino acids of human Type 1  $(\alpha_1)$  collagen (Figure 5) (SEQ. ID. NO. 2) placed behind the isopropyi- $\beta$ -D-thiogalactopyranoside (IPTG)-inducible tac promotor and also encoding β-lactamase is transformed into Escherichia coli proline auxotrophic strain NM519 (pro·) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 μg/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 5 mL of LB containing 100 50 μg/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37 °C, this culture is used to inoculate 1 L of M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100 μg/mL ampicillin, supplemented with all amino acids at 20 µg/mL except proline which is supplemented at 12.5 mg/L) in a 1.5 L shaker flask. After growth at 37 °C, 225 rpm, for 15-20 hours post-inoculation, the optical density at 600 nm is constant at approximately 0.5 OD/mL. The cells are harvested by centrifugation (5000 rpm, 5 minutes), the media decanted, and the cells resuspended in 1 L of 55. M9 minimal media containing 100 μg/mL ampicillin, 0.08M L-hydroxyproline, and 0.5M NaCl. Following growth for 1 hour at 37 °C, 225 rpm, IPTG is added to 1mM and the cultures allowed to grow for an additional 5-15 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. The hydroxyproline-containing collagen fragment is purified by ammonium sulfate fractionation and column chromatography.

# **EXAMPLE 10**

Construction and Expression in E. coli of the Human Collagen Type 1(a<sub>1</sub>) Gene with Optimized E. coli Codon Usage

### 5 A. Construction of the gene:

[0195] The nucleotide sequence of the helical region of human collagen Type I ( $\alpha_1$ ) gene flanked by 17 amino acids of the amino terminal extra-helical and 26 amino acids of the C-terminal extra-helical region is shown in Figure 27 (SEQ. ID. NO. 15). A tabulation of the codon frequency of this gene is given in Table I. The gene sequence shown in Figure 27 was first changed to reflect *E. coli* codon bias. An initiating methionine was inserted at the 5' end of the gene and a TAAT stop sequence at the 3' end. Unique restriction sites were identified or created approximately every 150 base pairs. The resulting gene (HUCOI<sup>EC</sup>, Figure 39A-39E) (SEQ. ID. NO. 20) has the codon usage given in Table II as shown below. Other sequences that approximate *E. coli* codon bias are also acceptable.

TABLE II

	Codon	Count	%age	Codon	Count	%age	Codon	Count	%age	Codon	Count	%age
	TTT-	6	0.56	тст-	3	0.28	TAT-	2	0.18	TGT-	0	0.00
	Phe			Ser			Tyr			Cys		
)	TTC-	9	0.85	TCC-	3	0.28	TAC-	2	0.18	TGC-	0	0.00
	Phe			Ser			Tyr			Cys		
	TTA-	0	0.00	TCA-	0	0.00	TAA-	0	0.00	TGA-***	0	0.00
	Leu			Ser			***					
	TTG-	0	0.00	TCG-	0	0.00	TAG-	0	0.00	TGG-	0	0.00
i	Leu			Ser			***			Trp		
	CTT-	0	0.00	ССТ-	13	1.22	CAT-	0	0.00	CGT-	26	2.45
	Leu			Pro			His			Arg		
	CTC-	1	0.09	CCC-	12	1.13	CAC-	3	0.28	CGC-	26	2.45
)	Leu			Pro			His			Arg		
	CTA-	1	0.09	CCA-	29	2.74	CAA-	5	0.47	CGA-	0	0.00
•	Leu			Pro			Gln			Arg		
	CTG-	19	1.79	CCG-	186	17.58	CAG-	25	2.36	CGG-	1	0.09
	Leu			Pro			Gln			Arg		
5 <sup>:</sup>	ATT-	3	0.28	ACT-	2	0.18	AAT-	0	0.00	AGT-	1	0.09
	lle			Thr			Asn			Ser		
	ATC-	4	0.37	ACC-	11	1.03	AAC-	11	1.03	AGC-	32	3.02
	lle		ŀ	Thr			Asn			Ser		
١٠	ATA-	0	0.00	ACA-	0	0.00	AAA-	38	3.59	AGA-	0	0.00
•	lle			Thr			Lys	_		Arg	_	
	ATG-	8	0.75	ACG-	4	0.37	AAG-	0	0.00	AGG-	0	0.00
	Met			Thr			Lys			Arg		
	GTT-	3	0.28	GCT-	10	0.94	GAT-	20	1.89	GGT-	148	13.98
i	Val			Ala			Asp			Gly		
	GTC-	5	0.47	GCC-	24	2.26	GAC-	14	1.32	GGC-	178	16.82
	Val		_	Ala			Asp			Gly	_	
	GTA-	0	0.00	GCA-	8	0.75	GAA-	40	3.78	CGA-	9	0.85
,	Val			Ala			Glu			Gly		
,	GTG-	12	1.13	GCG-	80	7.56	GAG-	9	0.85	GGG-	12	1.13
	Val		1	Ala			Glu		1	Gly		

[0196] Oligos of approximately 80 nucleotides were synthesized on a Beckman Oligo 1000 DNA synthesizer, cleaved and deprotected with aqueous  $NH_4OH$ , and purified by electrophoresis in 7M urea/12% polyacrylamide gels. Each set of oligos was designed to have an EcoR I restriction enzyme site at the 5' end, a unique restriction site near the 3' end, followed by the TAAT stop sequence and a Hind III restriction enzyme site at the very 3' end. The first four oligos, comprising the first 81 amino acids of the human collagen Type I ( $\alpha_1$ ) gene, are given in Figure 40 which shows the

sequence and restriction maps of synthetic oligos used to construct the first 243 base pairs of the human Type I ( $\alpha_1$ ) collagen gene with optimized *E. coli* codon usage. Oligos N1-1 (SEQ. ID. NO. 21) and N1-2 (SEQ. ID. NO. 22) were designed to insert an initiating methionine (ATG) codon at the 5' end of the gene.

[0197] In one instance, oligos N1-1 and N1-2 (1µg each) were annealed in 20 µL of T7 DNA polymerase buffer (40mM Tris·HC1 (pH 8.0), 5mM MgCl<sub>2</sub>, 5mM dithiothreitol, 50mM NaCl, 0.05 mg/mL bovine serum albumin) by heating at 90°C for 5 minutes followed by slow cooling to room temperature. After brief centrifugation at 14,000 rpm, 10 units of T7 DNA polymerase and 2 µL of a solution of all four dNTPs (dATP, dCTP, dCTP, dTTP, 2.5mM each) were added to the annealed oligos. Extension reactions were incubated at 37°C for 30 minutes and then heated at 70°C for 10 minutes. After cooling to room temperature, Hind III buffer (5 pL of 10x concentration), 20 μL of H<sub>2</sub>O, and 10 units of Hind III restriction enzyme were added and the tubes incubated at 37°C for 10 hours. Hind III buffer (2µL of 10x concentration), 13.5µL of 0.5M Tris·HC1 (pH 7.5), 1.8 µL of 1% Triton X100, 5.6 µL of H<sub>2</sub>O, and 20 U of EcoR I were added to each tube and incubation continued for 2 hours at 37°C. Digests were extracted once with an equal volume of phenol, once with phenol/chloroform/isoamyl alcohol, and once with chloroform/isoamyl alcohol. After ethanol precipitation, the pellet was resuspended in 10 µL of TE buffer (10mM Tris·HC1 (pH 8.0), 1mM EDTA). Resuspended pellet (4 µL) was ligated overnight at 16°C with agarose gel-purified EcoRI/Hind III digested pBSKS\* vector (1μg) using T4 DNA ligase (100 units). One half of the transformation mixture was transformed by heat shock into DH5α cells and 100 μL of the 1.0 mL transformation mixture was plated on Luria Broth (LB) agar plates containing 70 µg/mL ampicillin. Plates were incubated overnight at 37°C. Ampicillin resistant colonies (6-12) were picked and grown overnight in LB media containing 70 mg/mL ampicillin. Plasmid DNA was isolated from each culture by Wizard Minipreps (Promega Corporation, Madison WI) and screened for the presence of the approximately 120 base pair insert by digestion with EcoR I and Hind III and running the digestion products on agarose electrophoresis gels. Clones with inserts were confirmed by standard dideoxy termination DNA sequencing. The correct clone was named pBSN1-1 (Figure 41) and the collagen fragment has the nucleic acid sequence given in Figure 42 (SEQ. ID. NO. 25).

[0198] Oligos N1-3 (SEQ. ID. NO. 23) and N1-4 (SEQ. ID. NO. 24) (Figure 40) were synthesized, purified, annealed, extended, and cloned into pBSKS $^+$  following the same procedure given above for oligos N1-1 and N1-2. The resulting plasmid was named pBSN1-2A. To clone together the sections of the collagen gene from pBSN1-1 and pBSN1-2A, plasmid pBSN1-1 (1  $\mu$ g) was digested for 2 hours at 37°C with Rsr II and Hind III. The digested vector was purified by agarose gel electrophoresis. Plasmid pBSN1-2A (3  $\mu$ g) was digested for 2 hours at 37°C with Rsr II and Hind III and the insert purified by agarose gel electrophoresis. Rsr II/Hind III-digested pBSN1-1 was ligated with this insert overnight at 16°C with T4 DNA ligase. One half of the ligation mixture was transformed into DH5 $\alpha$  cells and 1/10 of the transformation mixture was plated on LB agar plates containing 70  $\mu$ g/mL ampicillin. After overnight incubation at 37°C, ampicillin-resistant clones were picked and screened for the presence of insert DNA as described above. Clones were confirmed by dideoxy termination sequencing. The correct clone was named pBSN1-2 (Figure 43) and the collagen fragment has the sequence given in Figure 44.

[0199] In similar manner, the remainder of the collagen gene is constructed such that the final DNA sequence is that given in Figure 39A-39E (SEQ. ID. NO. 19).

B) Expression of the gene in E. coli:

[0200] Following construction of the entire human collagen Type I (α<sub>1</sub>) gene with codon usage optimized for *E. coli*, the cloned gene is expressed in *E. coli*. A plasmid (pHuCol<sup>Ec</sup>, Figure 45) encoding the entire synthetic'collagen gene (Figure 39A-39E) placed behind the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible *tac* promotor and also encoding β-lactamase is transformed into *Escherichia coli* strain DH5α (*supE44* Δ/acU169 (φ80/acZ ΔM15) *hsd*R17 *rec*A1 endA1 gyrA96 *thi-*1 *rel*A1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 μg/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100 μg/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100 μg/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen is purified by ammonium sulfate fractionation and column chromatography. The yield is typically 15-25 mg/L of culture.

# **EXAMPLE 11**

. 5

10

15

20

55 Expression in E. coli of an 81 Amino Acid Fragment of Human Collagen Type I(α1) with Optimized E. coli Codon Usage

[0201] A plasmid (pTrcN1-2, Figure 46) encoding the gene sequence of the first 81 amino acids of human Type I ( $\alpha_1$ ) collagen with optimized *E. coli* codon usage cloned in fusion with a 6 histidine tag at the 5' end of the gene and

placed behind the isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG)-inducible trc promotor and also encoding  $\beta$ -lactarnase was constructed by subcloning the EcoR I/Hind III insert from pBSN1-2 into the EcoR I/Hind III site of plasmid pTrcB (Invitrogen, San Diego, CA). Plasmid pTrcN1-2 was transformed into Escherichia coli strain DH5α (supE44ΔlacU169 (φ80/ac/Z ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) by standard heat shock transformation. Transformation cultures were plated on Luria Broth (LB) containing 100 µg/mL ampicillin and after overnight growth a single ampicillinresistant colony was used to inoculate 5mL of LB containing 100 µg/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture was used to inoculate 50 mL of LB containing 100 μg/mL ampicillin in a 250 mL shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm was approximately 0.5 OD/mL. IPTG was added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells were harvested by centrifugation (5000 rpm, 10 minutes) and stored at -20°C. The 6 histidine tag-collagen fragment fusion was purified on nickel resin columns. Cell pellets were resuspended in 10 mL of 6M guanidine hydrochloride/ 20mM sodium phosphate/500mM Nacl (pH 7.8) and bound in two 5 mL batches to the nickel resin. Columns were washed two times with 4 mL of binding buffer (8M urea/20mM sodium phosphate/500mM NaCl (pH 7.8)), two times with wash buffer 1 (8M urea/20mM sodium phosphate/500mM NaCl (pH 6.0)), and two times with wash buffer 2 (8m urea/20mM sodium phosphate/500mM NaCl (pH 5.3). The 6 histidine tag-collagen fragment fusion was eluted from the column with 5mL of elution buffer (8M urea/20mM sodium phosphate/500mM NaCl (pH 4.0) in 1 mL fractions. Fractions were assessed for protein by gel electrophoresis and fusion-containing fractions were concentrated and stored at -20°C. The yield was typically 15-25 mg/L of culture.

[0202] The collagen is cleaved from the 6 histidine tag with enterokinase. Fusion-containing fractions are dialyzed against cleavage buffer (50mM Tris·HCI, pH 8.0/5mM CaCl<sub>2</sub>). After addition of enterokinase at 1  $\mu$ g enzyme for each 100  $\mu$ g fusion, the solution is incubated at 37°C for 4-10 hours. Progress of the cleavage is monitored by gel electrophoresis. The cleaved 6 histidine tag may be separated from the collagen fragment by passage over a nickel resin column as outlined above.

# **EXAMPLE 12**

., 5

20

25

30

40

Expression in E. coli of Fragments of Human Collagen Type I (a1) with Optimized E. coli Codon Usage

[0203] A plasmid (pN1-3, Figure 47) encoding the gene for the amino terminal 120 amino acids of human collagen Type I ( $\alpha_1$ ) with optimized *E. coli* codon usage placed behind the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible *tac* promotor and also encoding β-lactamase is transformed into *Escherichia coli* strain DH5α (sup E44 Δ/acU169 ( $\phi$ 80/acZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100  $\mu$ g/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100  $\mu$ g/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100  $\mu$ g/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen is purified by ammonium sulfate fractionation and column chromatography. The yield is typically 15-25 mg/L of culture.

# **EXAMPLE 13**

Expression in *E. coli* of a C-terminal Fragment of Human Collagen Type I (α<sub>1</sub>) with Optimized *E. coli* Codon Usage.

[0204] A plasmid (pD4, Figure 48) encoding the gene for the carboxy terminal 219 amino acids of human collagen Type I (α₁) with optimized *E. coli* codon usage placed behind the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible tac promotor and also encoding β-lactamase is transformed into *Escherichia coli* strain DH5α (sup E44 ΔlacU169 (φ80/acZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 μg/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100 μg/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100 μg/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rmp, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen fragment is purified by ammonium sulfate fractionation and column chromatography. The yield is typically 15-25 mg/L of culture.

# **EXAMPLE 14**

5

15

20

25

30

40

45

50

55

Construction and Expression in E. coli of the Human Collagen Type 1 (a2) Gene with Optimized E. coli Codon Usage

# A) Construction of the gene:

[0205] The nucleotide sequence of the helical region of human collagen Type I ( $\alpha_2$ ) gene flanked by 11 amino acids of the amino terminal extra-helical and 12 amino acids of the C-terminal extra-helical region is shown in Figures 49A-49E (SEQ. ID. NO. 29). A tabulation of the codon frequency of this gene is given in Table III below. The gene sequence shown in Figures 49A-49E was first changed to reflect E. coli codon bias. An initiating methionine was inserted at the 5' end of the gene and a TAAT stop sequence at the 3' end. Unique restriction sites are identified or created approximately every 150 base pairs. The resulting gene (HuCol(α<sub>2</sub>)<sup>Ec</sup>, Figures 50A-50E) (SEQ. ID. NO. 31) has the codon usage given in Table IV below. Other sequences that approximate E. coli codon bias are also acceptable.

Table III

						<u> </u>						
	Codon	Count	\$age	Codon	Count	8age	Codon	Count	sage	Codon	Count	\$age
	TTT-Phe	3	0.28	TCT-Ser	11	1.06	TAT-Tyr	2	0.19	TGT-Cys	0	0.00
	TTC-Phe	10	0.96	TCC-Ser	4	0.38	TAC-Tyr	3	0.28	TGC-Cys	0	0.00
	TTA-Leu	1	0.09	TCA-Ser	1	0.09	TAA-***	0	0.00	TGA-+++	0	0.00
	TTG-Leu	2	0.19	TCG-Ser	1	0.09	TAG-***	0	0.00	TGG-Trp	0	0.00
	CTT-Leu	16	1.54	CCT-Pro	125	12.06	CAT-His	7	0.67	CGT-Arg	17	1.64
	CTC-Leu	9	0.86	CCC-Pro	42	4.05	CAC-His	6	0.57	CGC-Arg	6	0.57
	CTA-Leu	2	0.19	CCA-Pro	30	2.89	CAA-Gln	13	1.25	CGA-Arg	6	0.57
	CTG-Leu	5	0.48	CCG-Pro	3	0.28	CAG-Gln	9	0.86	CGG-Arg	4	0.38
	ATT-Ile	14	1.35	ACT-Thr	14	1.35	AAT-Asn	10	0.96	AGT-Ser	11	1.06
	ATC-Ile	3	0.28	ACC-Thr	0	0.00	AAC-Aan	14	1.35	AGC-Ser	4	0.38
ĺ	ATA-Ile	1	0.09	ACA-Thr	3	0.28	AAA-Lys	15	1.44	AGA-Arg	16	1.54
	ATG-Met	5	0.48	ACG-Thr	1	0.09	AAG-Lys	16	1.54	AGG-Arg	6	0.57
	GTT-Val	20	1.93	GCT-Ala	82	7.91	GAT-Asp	20	1.93	GGT-Gly	179	17.27
	GTC-Val	5	0.48	GCC-Ala	3.7	1.64	GAC-Asp	5	0.48	GGC-Gly	74	7.14
	GTA-Val	3	0.28	GCA-Ala	9	0.86	GAA-Glu	29	2.79	GGA-Gly	80	7.72
	GTG-Val	10	0.96	GCG-Ala	0	0.00	GAG-Glu	16	1.54	GGG-Gly	16	1.54

26

Table IV

Codon	Count	\$age	Codon	Count	gage	nopoo	Count	ခွေအင်	Codon	Count	ваде
TTT-Phe	5	0.48	TCT-Ser	7	0.67	TAT-TYX	3	0.28	TGT-Cys	0	0.00
TTC-Phe	7	0.67	TCC-Ser	12	1.15	TAC-Tyr	2	0.19	TGC-Cys	0	0.00
TTA-Leu	0	0.00	TCA-Ser	0	0.00	TAA-***	0	0.00	TGA-***	0	0.00
TTG-Leu	0	0.00	TCG-Ser	0	0.00	TAG-***	0	0.00	TGG-Trp	်ဝ	0.00
CTT-Leu	1	0.09	CCT-Pro	10	0.96	CAT-His	2	0.19	CGT-Arg	37	3.55
CTC-Leu	1	0.09	CCC-Pro	0	0.00	CAC-His	11	1.05	CGC-Arg	18	1.72
CTA-Leu	0	0.00	CCA-Pro	15	1.44	CAA-Gln	7	0.67	CGA-Arg	0	0.00
CTG-Leu	32	3.07	CCG-Pro	177	17.00	CAG-Gln	15	1.44	CGG-Arg	0	0.00
ATT-Ile	11	1.05	ACT-Thr	3	0.28	AAT-Asn	6	0.57	AGT-Ser	0	0.00
ATC-Ile	7	0.67	ACC-Thr	6	0.57	AAC-Asn	18	1.72	AGC-Ser	13	1.24
ATA-Ile	0	0.00	ACA-Thr	0	0.00	AAA-Lys	25	2.40	AGA-Arg	0	0.00
ATG-Met	6	0.57	ACG-Thr	10	0.96	AAG-Lys	6	0.57	AGG-Arg	0	0.00
GTT-Val	18	1.72	GCT-Ala	30	2.88	GAT-Asp	11	1.05	GGT-Gly	209	20.07
GTC-Val	7	0.67	GCC-Ala	21	2.01	GAC-Asp	13	1.24	GGC-Gly	141	13.54
GTA-Val	9	0.85	GCA-Ala	20	1.92	GAA-Glu	33	3.17	GGA-Gly	0	0.00
GTG-Val	6	0.57	GCG-Ala	38	3.65	GAG-Glu	12	1.15	GGG-Gly	0	0.00

5

10.

**~15** 

20

25

35

40

45

[0206] Oligos of approximately 80 nucleotides are synthesized on a Beckman Oligo 1000 DNA synthesizer, cleaved and deprotected with aqueous  $NH_4OH$ , and purified by electrophoresis in 7M urea/12% polyacrylamide gels. Each set of oligos is designed to have an EcoR I restriction enzyme site at the 5' end, a unique restriction site near the 3' end; followed by the TAAT stop sequence and a Hind III restriction enzyme site at the very 3' end. Oligos  $N1-1(\alpha_2)$  and  $N1-2(\alpha_2)$  are designed to insert an initiating methionine (ATG) codon at the 5' end of the gene.

[0207] In one instance, oligos N1-1( $\alpha_2$ ) and N1-2( $\alpha_2$ ) (1  $\mu g$  each) (Figure 51 depicts sequence and restriction maps of synthetic oligos used to construct the first 240 base pairs of human Type I(α2) collagen gene with optimized E. coli codon usage) are annealed in 20 µL of T7 DNA polymerase buffer (40mM Tris·HCl (pH 8.0), 5mM MgCl<sub>2</sub>, 5mM dithiothreitol, 50mM NaCl, 0.05 mg/mL bovine serum albumin) by heating at 90°C for 5 minutes followed by slow cooling to room temperature. After brief centrifugation at 14,000 rpm, 10 units of T7 DNA polymerase and 2 μL of a solution of all four dNTPs (dATP, dGTP, dCTP, dTTP, 2.5mM each) are added to the annealed oligos. Extension reactions are incubated at 37°C for 30 minutes and then heated at 70°C for 10 minutes. After cooling to room temperature, Hind III buffer (5 μL of 10x concentration), 20 μL of H<sub>2</sub>O, and 10 units of Hind III restriction enzyme are added and the tubes incubated at 37°C for 10-16 hours. Hind III buffer (2 μL of 10x concentration), 13.5 μL of 0.5 Tris·HCl (pH 7.5); 1.8 μL of 1% Triton X100, 5.6 µL of H2O, and 20 U of EcoR I are added to each tube and incubation continued for 2 hours at 37°C. Digests are extracted once with an equal volume of phenol, once with phenol/chloroform/isoamyl alcohol, and once with chloroform/isoamyl alcohol. After ethanol precipitation, the pellet is resuspended in 10 µL of TE buffer (10mM Tris·HCI (pH 8.0), 1mM EDTA). Resuspended pellet (4 µL) is ligated overnight at 16°C with agarose gel-purified EcoRI/ Hind III digested pBSKS+ vector (1 µg) using T4 DNA ligase (100 units). One half of the transformation mixture is transformed by heat shock into DH5α cells and 100 μL of the 1.0 mL transformation mixture is plated on Luria Broth (LB) agar plates containing 70 μg/mL ampicillin. Plates are incubated overnight at 37°C. Ampicillin resistant colonies (6-12) are picked and grown overnight in LB media containing 70 μg/mL ampicillin. Plasmid DNA is isolated from each culture by Wizard Minipreps (Promega Corporation, Madison, WI) and screened for the presence of the approximately 120 base pair insert by digestion with EcoR I and Hind III and running the digestion products on agarose electrophoresis gels. Clones with inserts are confirmed by standard dideoxy termination DNA sequencing. The correct clone is named pBSN1-1( $\alpha_2$ ) Figure 52).

[0208] Oligos N1-3( $\alpha_2$ ) and N1-4( $\alpha_2$ ) are synthesized, purified, annealed, extended, and cloned into pBSKS+ following the same procedure given above for oligos N1-1( $\alpha_2$ ) and N1-2( $\alpha_2$ ). The resulting plasmid is named pBSN1-2A. To clone together the sections of the collagen gene from pBSN1-1( $\alpha_2$ ) (1  $\mu$ g) is digested for 2 hours at 37°C with BsrF I and Hind III. The digested vector is purified by agarose gel electrophoresis. Plasmid pBSn1-2( $\alpha_2$ ) (3  $\mu$ g) is digested for 2 hours at 37°C with BsrF I and Hind III and the insert purified by agarose gel electrophoresis. BsrF I/Hind III-digested pBSN1-1 is ligated with this insert overnight at 16°C with T4 DNA ligase. One half of the ligation mixture is transformed into DH5 $\alpha$  cells and 1/10 of the transformation mixture is plated on LB agar plates containing 70  $\mu$ g/mL ampicillin. After overnight incubation at 37°C, ampicillin-resistant clones are picked and screened for the presence of insert DNA as described above. Clones are confirmed by dideoxy termination sequencing. The correct clone is name

pBSN1-2( $\alpha_2$ ) (Figure 53) and the collagen fragment has the sequence given in Figure 54 (SEQ. ID. NO. 37). **[0209]** In a similar manner, the remainder of the collagen gene is constructed such that the final DNA sequence is that given in Figures 50A-50E (SEQ. ID. NO. 31).

B) Expression of the gene in E. coli:

5

10

15

20

25

30

45

55

[0210] Following construction of the entire human collagen Type I ( $\alpha$ 2) gene with codon usage optimized for *E. coli*, the cloned gene is expressed in *E. coli*. A plasmid (pHucol( $\alpha_2$ )<sup>Ec</sup>, Figure 55) encoding the entire synthetic collagen gene (Figures 50A-50E) placed behind the isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG)-inducible *tac* promotor and also encoding  $\beta$ -lactamase is transformed into *Escherichia coli* strain DH5 $\alpha$  ( $supE44 \Delta lacU169$  ( $\phi$ 80/ $lacZ \Delta M15$ ) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100  $\mu$ g/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100  $\mu$ g/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100  $\mu$ g/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen is purified by ammonium sulfate fractionation and column chromatography. The yield is typically 15-25 mg/L of culture.

**EXAMPLE 14A** 

Alternative Construction and Expression in *E. Coli* of the Human Collagen Type 1 (α2) Gene with Optimized *E. coli* Codon Usage

A) Construction of the gene:

[0211] The nucleotide sequence of the helical region of human collagen Type 1 ( $\alpha$ 2) gene flanked by 11 amino acids of the amino terminal extra-helical and 12 amino acids of the C-terminal extra-helical region is shown in Figures 49A-49E (SEQ. ID. NO. 29). A tabulation of the codon frequency of this gene is given in Table III. The gene sequence shown in Figures 49A-49E was first changed to reflect *E. coli* codon bias. An initiating methionine was inserted at the 5' end of the gene and a TAAT stop sequence at the 3' end. Unique restriction sites were identified or created at appropriate locations in the gene (approximately every 150 base pairs). The resulting gene (HuCol( $\alpha$ 2)<sup>Ec</sup>, Figures 50A-50E) (SEQ. ID. NO. 31) has the codon usage given in Table IV. Other sequences that approximate *E. coli* codon bias are also acceptable.

[0212] Oligonucleotides were synthesized on a Beckman Oligo 1000 DNA synthesizer, cleaved and deprotected with aqueous NH<sub>A</sub>OH, and purified by electrophoresis in 7M urea/12% polyacrylamide gels. Purified oligos (32.5 pmol) were dissolved in 20µL of ligation buffer (Boehringer Mannheim, Cat. No. 1635 379) and annealed by heating to 95°C followed by slow cooling to 20°C over 45 minutes. The annealed oligonucleotides were ligated for 5 minutes at room temperature with digested vector (1µg) using T4 DNA ligase (5 units). One half of the transformation mixture was transformed by heat shock into DH5 $\alpha$  cells and 100 $\mu$ L of the 1.0mL transformation mixture plated on Luria Broth (LB) agar plates containing 70µg/mL ampicillin. Plates were incubated overnight at 37°C. Ampicillin resistant colonies (6-12) were picked and grown overnight in LB media containing 70µg/mL ampicillin. Plasmid DNA was isolated from each culture by QIAprep Miniprep (Qiagen, Valencia, CA) and screened for the presence of insert by digestion with flanking restriction enzymes and running the digestion products on agarose electrophoresis gels. Clones with inserts were confirmed by standard dideoxy termination DNA sequencing. To clone together the sections of the collagen gene, and insert covering a flanking portion of the gene was ligated into vector containing the neighboring gene portion. Inserts were isolated from plasmids and vectors were cut by double digestion for 2 hours at 37°C with the appropriate restriction enzymes. The digested vector and insert were purified by agarose gel electrophoresis. Insert and vector were ligated for 5 minutes at room temperature following the procedure in the Rapid DNA Ligation Kit (Boehringer Mannheim). One half of the ligation mixture is transformed into DH5α cells and 1/10 of the transformation mixture was plated on LB agar plates containing 70µg/mL ampicillin. After overnight incubation at 37°C, ampicillin-resistant clones were picked and screened for the presence of insert DNA as described above. Clones were confirmed by dideoxy termination sequencing.

[0213] In a similar manner, the remainder of the collagen gene was constructed such that the final DNA sequence is that given in Figures 50A-50E (SEQ. ID. NO. 31).

B) Expression of the gene in E. coli:

[0214] Following construction of the entire human collagen Type 1( $\alpha$ 2) gene with codon usage optimized for *E. coli*, the cloned gene is expressed in *E. coli*. A plasmid (pHuCol)( $\alpha$ 2)<sup>Ec</sup>, Figure 55) encoding the entire collagen gene (Figures 50A-50E) placed behind the isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG)-inducible *tac* promoter and also encoding  $\beta$ -lactamase is transformed into *Escherichia coli* strain DH5 $\alpha$  (sup*E*44  $\Delta$ 1acU169 ( $\phi$ 80/acZ  $\Delta$ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 $\mu$ g/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100 $\mu$ g/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100 $\mu$ g/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen is purified by ammonium sulfate fractionation and column chromatograph. The yield is typically 15-25 mg/L of culture.

**EXAMPLE 15** 

5

10

15

45

50

55

Expression in E. coli of Fragments of Human Collagen Type I(α2) with Optimized E. coli Codon Usage

[0215] A plasmid (pN1-2, Figure 56) encoding the gene for the amino terminal 80 amino acids of human collagen Type I(α<sub>2</sub>) (SEQ. ID. NO. 31, Fig. 54) with optimized *E. coli* codon usage placed behind the isopropyl-β-D-thiogalact-opyranoside (IPTG)-inducible *tac* promotor and also encoding (β-lactamase is transformed into *Escherichia coli* strain DH5α (*sup*E44 Δ/*ac*U169 (φ80/*ac*Z ΔM15) *hsd*R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 μg/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100 μg/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100 μg/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen is purified by ammonium sulfate fractionation and column chromatography. The yield is typically 15-25 mg/L of culture.

# **EXAMPLE 16**

35 Hydroxyproline Incorporation Into Proteins In E. coli Under Proline Starvation Conditions

[0216] Seven plasmids, pGEX-4T.1 (Fig. 73), pTrc-TGF (Fig. 74), pMal-C2 (Fig. 1), pTrc-FN (Fig. 75), pTrc-FN-TGF (Fig. 76), pTrc-FN-Bmp (Fig. 77) and pGEX-HuColl<sup>Ec</sup>, each separately containing genes encoding the following proteins: glutathione S-transferase (GST), the mature human TGF-β1 polypeptide (TGF-β1), mannose-binding protein (MBP), a 70 kDA fragment of human fibronectin (FN), a fusion of FN and TGF-β1 (FN-TGF-β1), a fusion of FN and human bone morphogenic protein 2A (FN-BMP-2A), and a fusion of GST and collagen (GST-Coll), were used individually to transform proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 100 μg/ml ampicillin. After overnight incubation at 37°C, a single colony from a fresh transformation plate was used to inoculate 5 ml of LB media containing 400 mg ampicillin. After overnight growth at 37°C, this culture was centrifuged, the supernatant discarded, and the cell pellet washed twice with 5 ml of M9 medium (1X M9 salts, 0.5% glucose, 1 mM MgCl<sub>2</sub>, 0.01% thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 400 µg/ml ampicillin). The cells were finally resuspended in 5 ml of M9 medium. After incubation with shaking at 37°C for 30 minutes, trans-4-hydroxyproline was added to 40mM, NaCl to 0.5 M, and isopropyl-B-D-thiogalactopyyranoside to 1.5 mM. In certain cultures one of these additions was not made, as indicated in the labels for the lanes of the gels. After addition, incubation with shaking at 37°C was continued. After 4 hours, the cultures were centrifuged, the supernatants discarded, and the cell pellets resuspended in SDS-PAGE sample buffer (300 mM Tris (pH6.8)/0.5% SDS/10% glycerol/0.4M β-mercapthoethanol/0.2% bromophenol blue) to 15 OD600nm AU/ml, placed in boiling water bath for five minutes, and electrophoresed in denaturing polyacrylaminde gels. Proteins in the gels were visualized by staining with Coomassie Blue R250. The results of the gels are depicted in scans shown in Figs. 57-59. The scans relating to GST, TGF-β1, MBP, FN, FN-TGF-β1, and FN-BMP-2A (Figs. 57 and 58) show three lanes relating to each peptide, i.e., one lane indicating +NaCl/+Hyp wherein NaCl (hyperosmotic) and trans-4-hydroxyproline are present; one lane indicating -NaCl wherein trans-4-hydroxyproline is present but NaCl is not; and one lane indicating -Hyp which is +NaCl but absent trans-4-hydroxyproline. Asterisks on the scans mark protein bands which correspond

to the expressed target protein. The instances in which target protein was expressed all involve +NaCl in connection with +Hyp thus demonstrating +NaCl and +Hyp dependence.

[0217] The scan shown in Fig. 59 relating to GST-collagen shows four lanes relating to GST-Coll, i.e., one lane indicating +Hyp/+NaCl/-IPTG wherein *trans*-4-hydroxyproline and NaCl are present but IPTG (the protein expression inducer) is not and since there is no inducer, there is no target protein band; one lane indicating +NaCl/+IPTG/-Hyp wherein NaCl and IPTG are present but *trans*-4-hydroxyproline is not and, since *trans*-4-hydroxyproline is not present no target protein band is evident; one lane indicating +NaCl/+Pro/+IPTG wherein NaCl, proline and IPTG are present, but since the target protein is not stable when it contains proline, there is no target protein band; and one lane designated +IPTG/+NaCl/+Hyp wherein IPTG, NaCl and *trans*-4-hydroxyproline are present and since the protein is stabilized by the presence *of trans*-4-hydroxyproline an asterisk marked protein band is evident.

# **EXAMPLE 17**

5

15

20

25

30

35

40

45

50:

55

Hydroxyproline incorporation into a collagen-like peptide in E. coli.

[0218] A plasmid (pGST-CM4, Figure 60) containing the gene for collagen mimetic 4 (CM4, Figure 61) (SEQ. ID. NO. 39) genetically linked to the 3' end of the gene for S. japonicum glutathione S-transferase was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 100 μg/ml ampicillin. After overnight incubation at 37° C, a single colony from a fresh transformation plate was used to inoculate 5 ml of LB media containing 100 μg/ml ampicillin. After overnight growth at 37° C, 500 μl of this culture was centrifuged, the supernatent discarded, and the cell pellet washed once with 500 µl of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl<sub>2</sub>, 0.01 % thiamine, 200 µg/ml glycine, 200µg/ml alanine, 100 µg/ml of the other amino acids except proline, and 400 µg/ml ampicillin). The cells were finally suspended in 5 ml of M9 medium containing 10 μg/ml proline and 2 ml of this was used to inoculate 30 ml of M9 medium containing 10 μg/ml proline. After incubation with shaking at 37° C for 8 hours, the culture was centrifuged and the cell pellet washed once with M9 medium containing 5 μg/ml proline. The pellet was resuspended in 15 ml of M9 medium containing 5 μg/ml of proline and this culture was used to inoculate 1 L of M9 medium containing 5 μg/ml of proline. This culture was grown for 18 hours at 37° C to proline starvation. At this time, the culture was centrifuged, the cells washed once with M9 medium (with no proline), and the cells resuspended in 1 L of M9 medium containing 80 mM hydroxyproline, 0.5 M NaCl, and 1.5 mM isopropylβ-D-thiogalactopyranoside. Incubation was continued at 37° C with shaking for 22 hours. The cultures were centrifuged and the cell pellets stored at -20°C until processed further.

# **EXAMPLE 18**

Proline incorporation into a collagen-like peptide in E. coli.

[0219] A plasmid (pGST-CM4, Figure 60) containing the gene for collagen mimetic 4 (CM4, Figure 61) (SEQ. ID. NO. 39) genetically linked to the 3' end of the gene for S. japonicum glutathione S-transferase was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 100 µg/ml ampicillin. After overnight incubation at 37° C, a single colony from a fresh transformation plate was used to inoculate 5 ml of LB media containing 100 μg/ml ampicillin. After overnight growth at 37° C, 500 μl of this culture was centrifuged, the supernatent discarded, and the cell pellet washed once with 500 µl of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl<sub>2</sub>, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 400 µg/mL ampicillin). The cells were finally resuspended in 5 ml of M9 medium containing 10 μg/ml proline and 2 ml of this was used to inoculate 30 ml of M9 medium containing 10 μg/ml proline. This culture was incubated with shaking at 37° C for 8 hours. The culture was centrifuged and the cell pellet washed once with M9 medium containing 5 μg/ml proline. The pellet was resuspended in 15 ml of M9 medium containing 5 μg/ml of proline and this culture was used to inoculate 1 L of M9 medium containing 5 µg/ml of proline. This culture was grown for 18 hours at 37°C to proline starvation. At this time, the culture was centrifuged, the cells washed once with M9 medium (with no proline), and finally the cells were resuspended in 1 L of M9 medium containing 2.5 mM proline, 0.5 M NaCl, and 1.5 mM isopropyl-p-β-thiogalactopyranoside. Incubation was continued at 37° C with shaking for 22 hours. The cultures were then centrifuged and the cell pellets stored at -20°C until processed further.

# **EXAMPLE 19**

Purification of hydroxyproline-containing collagen-like peptide from E. coli

[0220] The cell pellet from a 1 L fermentation culture prepared as described in Example 17 above, was resuspended

in 20 ml of Dulbecco's phosphate buffered saline (pH 7.1) (PBS) containing 1 mM EDTA, 100 μM PMSF, 0.5 μg/ml E64, and 0.7 μg/ml pepstatin (resuspension buffer). The cells were lysed by twice passing through a French press. Following lysis, the suspension was centrifuged for 30 minutes at 30,000 xg. The supernatent was discarded and the pellet washed once with 5 ml of resuspension buffer containing 1 M urea and 0.5% Triton X100 followed by one wash with 7 ml of resuspension buffer without urea or Triton X100. The pellet was finally resuspended in 5 ml of 6M guanidine hydrochloride in Dulbecco's phosphate buffered saline (pH7.1) containing 1 mM EDTA and 2 mM β-mercaptoethanol and sonicated on ice for 3 x 60 seconds (microtip, power = 3.5, Heat Systems XL-2020 model sonicator). The sonicated suspension was incubated at 4° C for 18 hours and then centrifuged at 14,000 rpm in a microcentrifuge. The supernatent (6 ml) was dialyzed (10,000 MWCO) against 4 x 4 L of distilled water at 4°C. The contents of the dialysis tubing were transferred to a 150 ml round bottom flask and lyophilized to dryness. The residue (~30 mg) was dissolved in 3 ml of 70% formic acid and 40 mg of cyanogen bromide was added. The flask was flushed once with nitrogen, evacuated, and allowed to stir for 18 hours at room temperature. The contents of the flask were taken to dryness in vacuo at room temperature, the residue resuspended in 5 ml of distilled water and evaporated to dryness again. This was repeated 2 times. The residue was finally dissolved in 2 ml of 0.2% trifluoroacetic acid (TFA). The trifluoroacetic acid-soluble material was applied in 100 μl aliquots to a Poros R2 column (4.6 mm x 100 mm) running at 5 ml/min. with a starting buffer of 98% 0.1% trifluoroacetic acid in water/2% 0.1 % TFA in acetonitrile. The hydroxyproline-containing protein was eluted with of gradient of 2% 0.1% TFA/acetonitrile to 40% 0.1% TFA/acetonitrile over 25 column volumes (Fig. 62A). The collagen-mimetic eluted between 18 and 23% 0.1% TFA/acetonitrile. Figure 62A is a chromatogram of the elution of hydroxyproline containing CM4 from a Poros RP2 column (available from Perseptive Biosystems, Framingham, MA). The arrow indicates the peak containing hydroxyproline containing CM4. Fractions were assayed by SDS-PAGE and collagen mimetic-containing fractions were pooled and lyophilized. Lyophilized material was stored at -20°

# **EXAMPLE 20**

5

۵

10

15

20

25

30

35

45

50

Purification of proline-containing collagen-like peptide from E. coli

[0221] The cell pellet from a 500 ml fermentation culture prepared as described in Example 18 above, was resuspended in 20 ml of Dulbecco's phosphate buffered saline (pH 7.1) (PBS) containing 10 mM EDTA, 100 µM PMSF, 0.5 μg/ml E64, and 0.06 μg/ml aprotinin. Lysozyme (2 mg) was added and the suspension incubated at 4° C for 60 minutes. The suspension was sonicated for 5 x 60 seconds (microtip, power = 3.5, Heat Systems XL-2020 model sonicator). The sonicated suspension was centrifuged at 20,000 xg for 15 minutes. The supernatent was adjusted to 1% Triton X100 and incubated for 30 minutes at room temperature with 7 ml of glutathione sepharose 4B pre-equilibrated in PBS. The suspension was centrifuged at 500 rpm for 3 minutes. The supernatent decanted, and the resin washed 3 times with 8 ml of PBS. Bound proteins were eluted with 3 aliquots (2 ml each, 10 minutes gentle rocking at room temperature) of 10 mM glutathione in 50 mM Tris (pH 8.0). Eluants were combined and dialyzed (10,000 MWCO) against 3 x 4 L of distilled water at 4° C. The contents of the dialysis tubing were transferred to a 150 ml round bottom flask and lyophilized to dryness. The residue was dissolved in 3 ml of 70% formic acid and 4 mg of cyanogen bromide was added. The flask was flushed once with nitrogen evacuated, and allowed to stir for 18 hours at room temperature. The contents of the flask were taken to dryness in vacuo at room temperature, the residue resuspended in 5 ml of distilled water, and evaporated to dryness again. This was repeated 2 times. The residue was finally dissolved in 2 ml of 0.2% trifluoroacetic acid (TFA). The trifluoroacetic acid-soluble material was applied in 100 μl aliquots to a Poros R2 column (4.6 mm x 100 mm) running at 5 ml/min. with a starting buffer of 98% 0.1% trifluoroacetic acid in water/2% 0.1% TFA in acetonitrile. Bound protein was eluted with of gradient of 2% 0.1% TFA/acetonitrile to 40% 0.1% TFA/acetonitrile over 25 column volumes (Figure 62B). The collagen-mimetic eluted between 24 and 27% 0.1% TFA/acetonitrile. Figure 62B is a chromatogram of the elution of proline containing CM4 from a Poros RP2 column. The arrow indicates the peak containing proline containing CM4. Fractions were assayed by SDS-PAGE and collagen mimetic-containing fractions were pooled and lyophilized. Lyophilized material was stored at -20° C.

# **EXAMPLE 21**

Amino acid analysis of hydroxyproline-containing collagen mimetic and proline-containing collagen mimetic.

[0222] Approximately 30 μg of purified hydroxyproline-containing collagen mimetic and proline-containing collagen mimetic prepared as described in Examples 19 and 20, respectively, were dissolved in 250 μl of 6N hydrochloric acid in glass ampules. The ampules were flushed two times with nitrogen, sealed under vacuum, and incubated at 110°C for 23 hours. Following hydrolysis, samples were removed from the ampules and taken to dryness in vacuo. The samples were dissolved in 15 μl of 0.1N hydrochloric acid and subjected to amino acid analysis on a Hewlett Packard

AminoQuant 1090 amino acid analyzer utilizing standard OPA and FMOC derivitization chemistry. Examples of the results of the amino acid analysis that illustrate the region of the chromatograms where the secondary amino acids (proline and hydroxyproline) elute are shown in Figures 63A through 63D. These Figures also show chromatograms of proline and hydroxyproline amino acid standards. More particularly, Figure 63A, depicts a chromatogram of a proline amino acid standard (250 pmol). \*indicates a contaminating peak; Figure 63B depicts a chromatogram of a hydroxyproline amino acid standard (250 pool). \*indicates a contaminating peak. Figure 63C depicts an amino analysis chromatogram of the hydrolysis of proline-containing CM4. Only the region of the chromatogram where proline and hydroxyproline elute is shown. \*indicates a contaminating peak. Figure 63D depicts an amino acid analysis chromatogram of the hydrolysis of hydroxyproline-containing CM4. Only the region of the chromatogram where proline and hydroxyproline elute is shown. \*indicates a contaminating peak.

# **EXAMPLE 22**

5

10

15

25

30

35

40

45

55

Determination of proline starvation conditions for E. coli (strain JM109 (F-))

[0223] A plasmid (pGST-CM4, Figure 60) containing the gene for collagen mimetic 4 (CM4, Figure 61) genetically linked to the 3' end of the gene for S. japonicum glutathione S-transferase was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 100 µg/ml ampicillin. After overnight incubation at 37 °C, a single colony from a fresh transformation plate was used to inoculate 2 ml of M9 media (1X M9 salts, 0.5 % glucose, 1 mM MgCl<sub>2</sub>, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 200 μg/ml carbenicillin) and containing 20 μg/ml proline. After growth at 37° C with shaking for 8 hours, 1.5 ml was used to inoculate 27 ml of M9 media containing 45 µg/ml proline. After incubation at 37° C with shaking for 7 hours, the culture was centrifuged, the cell pellet washed with 7 ml of M9 media with no proline, and finally resuspended in 17 ml of M9 media with no proline. This culture was used to inoculate four 35 ml cultures of M9 media containing 4 µg/ml proline at an OD600 of 0.028. Cultures were incubated with shaking at 37° C and the OD600 monitored. After 13.5 hours growth, the OD600 had plateaued. At this time, one culture was supplemented with proline at 15 μg/ml, one with hydroxyproline at 15 μg/ml, one with all of the amino acids at 15 μg/ ml except proline and hydroxyproline, and one culture with nothing. Incubation was continued and the OD600 monitored for a total of 24 hours. Figure 64 is a graph of OD600 vs. time for cultures of JM109 (F-) grown to plateau and then supplemented with various amino acids. The point at which the cultures were supplemented is indicated with an arrow. Proline starvation is evident since only the culture supplemented with proline continued to grow past plateau.

# **EXAMPLE 23**

Hydroxyproline Incorporation Into Type I (α1) Collagen in E. coli

[0224] A plasmid (pHuCol( $\alpha$ 1)<sup>Ec</sup>, Figure 65) containing the gene for Type I ( $\alpha$ 1) collagen with optimized E. coli codon usage (Figure 39A-39E) (SEQ. ID. NO. 19) under control of the tac promoter and containing the gene for chloramphenicol resistance was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 20 μg/ml chloramphenicol. After overnight incubation at 37 °C, a single colony from a fresh transformation plate was used to inoculate 100 ml of LB media containing 20 μg/ml chloramphenicol. This culture was grown to an OD600nm of 0.5 and 100 μl aliquots transferred to 1.5 ml tubes. The tubes were stored at -80 ° C. For expression, a tube was thawed on ice and used to inoculate 25 ml of LB media containing 20 μg/ml chloramphenicol. After overnight growth at 37° C, a four ml aliquot was withdrawn, centrifuged, the cell pellet washed once with 1 ml of 2x YT media containing 20 μg/ml chloramphenicol, and the washed cells used to inoculate 1 L of 2x YT medium containing 20 μg/ml chloramphenicol. This culture was grown at 37° C to an OD600nm of 0.8. The culture was centrifuged and the cell pellet washed once with 100 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl<sub>2</sub>, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 20 µg/ml chloramphenicol). The cells were resuspended in 910 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl<sub>2</sub>, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 20 µg/ml chloramphenicol) and allowed to grow at 37° C for 30 minutes. NaCl (80 ml of 5 M), hydroxyproline (7.5 ml of 2M), and IPTG (500 µl of 1 M) were added and growth continued for 3 hours. Cells were harvested by centrifugation and stored at -20° C.

# **EXAMPLE 24**

5

15

20

30

40

50

Hydroxyproline Incorporation Into Type I ( $\alpha$ 2) in E. coli

[0225] A plasmid (pHuCol( $\alpha$ 2)<sup>Ec</sup>, Figure 66) containing the gene for Type I ( $\alpha$ 2) collagen with optimized E. coli codon usage (Figure 50A-50E) (SEQ. ID. NO. 31) under control of the tac promoter and containing the gene for chloramphenicol resistance was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 20 µg/ml chloramphenicol. After overnight incubation at 37° C, a single colony from a fresh transformation plate was used to inoculate 100 ml of LB media containing 20 µg/ml chloramphenicol. This culture was grown to an OD600nm of 0.5 and 100 µl aliquots transferred to 1.5 ml tubes. The tubes were stored at -80 ° C. For expression, a tube was thawed on ice and used to inoculate 25 ml of LB media containing 20 μg/ml chloramphenicol. After overnight growth at 37° C, a four ml aliquot was withdrawn, centrifuged, the cell pellet washed once with 1 ml of 2x YT media containing 20 µg/ml chloramphenicol, and the washed cells used to inoculate 1 L of 2x YT medium containing 20 µg/ml chloramphenicol. This culture was grown at 37° C to an OD600nm of 0.8. The culture was centrifuged and the cell pellet washed once with 100 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl<sub>2</sub>, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 20 µg/ml chloramphenicol). The cells were resuspended in 910 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl<sub>2</sub>, 0.01 % thiamine, 200 µg/ml glycine, 200 µg/ml alanine, 100 µg/ml of the other amino acids except proline, and 20 µg/ml chloramphenicol) and allowed to grow at 37° C for 30 minutes. NaCl (80 ml of 5 M), hydroxyproline (7.5 ml of 2M), and IPTG (500 µl of 1 M) were added and growth continued for 3 hours. Cells were harvested by centrifugation and stored at -20° C.

# **EXAMPLE 25**

25 Hydroxyproline Incorporation Into a C-terminal Fragment of Type I (α1) Collagen in E. coli

[0226] A plasmid (pD4-α1, Figure 67) encoding the gene for the carboxy terminal 219 amino acids of human Type I (α1) collagen with optimized E. coli codon usage fused to the 3'-end of the gene for glutathione S-transferase and under control of the tac promoter and containing the gene for ampicillin resistance was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 100 μg/ml ampicillin. After overnight incubation at 37° C, a single colony from a fresh transformation plate was used to inoculate 100 ml of LB media containing 100 µg/ml ampicillin. This culture was grown to an OD600nm of 0.5 and 100 μl aliquots transferred to 1.5 ml tubes. The tubes were stored at -80° C. For expression, a tube was thawed on ice and used to inoculate 25 ml of LB media containing 400 µg/ml ampicillin. After overnight growth at 37° C, a four ml aliquot was withdrawn, centrifuged, the cell pellet washed once with 1 ml of 2x YT media containing 400 µg/ml ampicillin, and the washed cells used to inoculate 1 L of 2x YT medium containing 400 μg/ml ampicillin. This culture was grown at 37° C to an OD600nm of 0.8. The culture was centrifuged and the cell pellet washed once with 100 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl $_2$ , 0.01 % thiamine, 200  $\mu$ g/ml glycine, 200  $\mu$ g/ml alanine, 100  $\mu$ g/ml of the other amino acids except proline, and 400 µg/ml ampicillin). The cells were resuspended in 910 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl<sub>2</sub>, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 400 µg/ml ampicillin) and allowed to grow at 37° C for 30 minutes. NaCl (80 ml of 5 M), hydroxyproline (7.5 ml of 2M), and IPTG (500 pl of 1 M) were added and growth continued for 3 hours. Cells were harvested by centrifugation and stored at -20° C.

# 45 EXAMPLE 26

Hydroxyproline Incorporation Into a C-terminal Fragment of Type I (α2) Collagen in E. coli

[0227] A plasmid (pD4- $\alpha$ 2, Figure 68) encoding the gene for the carboxy terminal 219 amino acids of human Type I ( $\alpha$ 2) collagen with optimized *E. coli* codon usage as constructed in accordance with Example 14A fused to the 3'-end of the gene for glutathione *S*-transferase and under control of the *tac* promoter and containing the gene for ampicillin resistance was used to transform by electroporation proline auxotrophic *E. coli* strain JM109 (F-). Transformation cultures were plated on LB agar containing 100  $\mu$ g/ml ampicillin. After overnight incubation at 37° C, a single colony from a fresh transformation plate was used to inoculate 100 ml of LB media containing 100  $\mu$ g/ml ampicillin. This culture was grown to an OD600nm of 0.5 and 100  $\mu$ l aliquots transferred to 1.5 ml tubes. The tubes were stored at -80° C. For expression, a tube was thawed on ice and used to inoculate 25 ml of LB media containing 400  $\mu$ g/ml ampicillin. After overnight growth at 37° C, a four ml aliquot was withdrawn, centrifuged, the cell pellet washed once with 1 ml of 2x YT media containing 400  $\mu$ g/ml ampicillin, and the washed cells used to inoculate 1 L of 2x YT medium containing

400  $\mu$ g/ml ampicillin. This culture was grown at 37° C to an OD600nm of 0.8. The culture was centrifuged and the cell pellet washed once with 100 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl<sub>2</sub>, 0.01 % thiamine, 200  $\mu$ g/ml glycine, 200  $\mu$ g/ml alanine, 100  $\mu$ g/ml of the other amino acids except proline, and 400  $\mu$ g/ml ampicillin). The cells were resuspended in 910 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl<sub>2</sub>, 0.01 % thiamine, 200  $\mu$ g/ml glycine, 200  $\mu$ g/ml alanine, 100  $\mu$ g/ml of the other amino acids except proline, and 400  $\mu$ g/ml ampicillin) and allowed to grow at 37° C for 30 minutes. NaCl (80 ml of 5 M), hydroxyproline (7.5 ml of 2M), and IPTG (500  $\mu$ l of 1 M) were added and growth continued for 3 hours. Cells were harvested by centrifugation and stored at -20° C.

# **EXAMPLE 27**

10

15

20

25

40

45

50

55

Purification of Hydroxyproline-containing C-terminal Fragment of Type I ( $\alpha$ 1) Collagen

[0228] Cell paste harvested from a 1 L culture grown as in Example 25 was resuspended in 30 ml of lysis buffer (2M urea, 137mM NaCl, 2.7mM KCl, 4.3mM Na<sub>2</sub>HPO<sub>4</sub>, 1.4mM KH<sub>2</sub>PO<sub>4</sub>, 10mM EDTA, 10mM βME, 0.1% Triton X-100, pH 7.4) at 4°C. Lysozyme (chicken egg white) was added to 100 μg/ml and the solution incubated at 4 °C for 30 minutes. The solution was passed twice through a cell disruption press (SLM Instruments, Rochester, NY) and then centrifuged at 30,000 x g for 30 minutes. The pellet was resuspended in 30 ml of 50 mM Tris-HCl, pH 7.6, centrifuged at 30,000 x g for 30 minutes, and the pellet solubilized in 25 ml of solubilization buffer (8M urea, 137mM NaCl, 2.7mM KCl, 4.3mM Na<sub>2</sub>HPO<sub>4</sub>, 1.4mM KH<sub>2</sub>PO<sub>4</sub>, 5mM EDTA, 5mM βME). The solution was centrifuged at 30,000xg for 30 minutes and supernatent dialyzed against two changes of 4 L of distilled water at 4°C. Following dialysis, the entire mixture was lyophilized. The lyophilized solid was dissolved in 0.1M HCl in a flask with stirring. After addition of a 5-fold excess of crystalline BrCN, the flask was evacuated and filled with nitrogen. Cleavage was allowed to proceed for 24 hrs, at which time the solvent was removed in vacuo. The residue was dissolved in 0.1% trifluoroacetic acid (TFA) and purified by reverse-phase HPLC using a Vydac C4 RP-HPLC column (10x250mm, 5μ, 300 Å) on a BioCad Sprint system (Perceptive Biosystems, Framingham, MA). Hydroxyproline-containing D4 protein was eluted with a gradient of 15-40% acetonitrile/0.1% TFA.

### **EXAMPLE 28**

30 Purification of Hydroxyproline-containing C-terminal Fragment of Type I (α2) Collagen

[0229] Cell paste harvested from a 1 L culture grown as in Example 26 was resuspended in 30 ml of lysis buffer (2M urea, 137mM NaCl, 2.7mM KCl, 4.3mM Na $_2$ HPO $_4$ , 1.4mM KH $_2$ PO $_4$ , 10mM EDTA, 10mM βME, 0.1% Triton X-100, pH 7.4) at 4°C. Lysozyme (chicken egg white) was added to 100 μg/ml and the solution incubated at 4°C for 30 minutes. The solution was passed twice through a cell disruption press (SLM Instruments, Rochester, NY) and then centrifuged at 30,000 x g for 30 minutes. The pellet was resuspended in 30 ml of 50 mM Tris-HCl, pH 7.6, centrifuged at 30,000 x g for 30 minutes, and the pellet solubilized in 25 ml of solubilization buffer (8M urea, 137mM NaCl, 2.7mM KCl, 4.3mM Na $_2$ HPO $_4$ , 1.4mM KH $_2$ PO $_4$ , 5mM EDTA, 5mM βME). The solution was centrifuged at 30,000xg for 30 minutes and supernatent dialyzed against two changes of 4 L of distilled water at 4°C. Following dialysis, the entire mixture was lyophilized. The lyophilized solid was dissolved in 0.1 M HCl in a flask with stirring. After addition of a 5-fold excess of crystalline BrCN, the flask was evacuated and filled with nitrogen. Cleavage was allowed to proceed for 24 hrs, at which time the solvent was removed in vacuo. The residue was dissolved in 0.1% trifluoroacetic acid (TFA) and purified by reverse-phase HPLC using a Vydac C4 RP-HPLC column (10x250mm, 5μ, 300 Å) on a BioCad Sprint system (Perceptive Biosystems, Framingham, MA). Hydroxyproline-containing D4 protein was eluted with a gradient of 15-40% acetonitrile/0.1 % TFA over a 45 minute period. Protein D4- $\alpha$ 2 eluted at 25% acetonitrile/0.1 % TFA.

# **EXAMPLE 29**

Amino Acid Composition Analysis of Hydroxyproline-containing C-terminal Fragment of Type I (α1) Collagen

[0230] Protein D4-α1 (10μg) purified as in Example 27 was taken to dryness in vacuo in a 1.5 ml microcentrifuge tube. A sample was subjected to amino acid analysis at the W.M. Keck Foundation Biotechnology Resource Laboratory (New Haven, CT) on an Applied Biosystems sequencer equipped with an on-line HPLC system. The experimentally determined sequence of the first 13 amino acids (SEQ. ID. NO. 41) and the sequence predicted from the DNA sequence (SEQ. ID. NO. 42) are shown in Figure 69. A sample of protein D4-al was subjected to mass spectral analysis on a VG Biotech BIO-Q quadrople analyzer at M-Scan, Inc. (West Chester, PA). The mass spectrum and the predicted molecular weight of protein D4-α1 if it contained 100% hydroxyproline in lieu of proline are given in Figure 70. The predicted molecular weight of protein D4-α1 containing 100% hydroxyproline in lieu of proline is 20807.8 Da. The

experimentally determined molecular weight was 20807.5 Da.

## **EXAMPLE 30**

10

15

20

30

40

45

5 Construction of Carboxy Terminal 219 Amino Acids of Human Collagen Type I (α1) Fragment Gene with Optimized E. Coli Codon Usage.

[0231] The nucleotide sequence of the 657 nucleotide gene for the carboxy terminal 219 amino acids of human Type I ( $\alpha$ 1) collagen with optimized *E. Coli* codon usage is shown in Figure 71. For synthesis of this gene, unique restriction sites were identified or created approximately every 150 base pairs. Oligos of approximately 80 nucleotides were synthesized on a Beckman Oligo 1000 DNA synthesizer, cleaved and deprotected with aqueous NH<sub>4</sub>OH, and purified by electrophoresis in 7M urea/12% polyacrylamide gels. Each set of oligos was designed to have an EcoR I restriction enzyme site at the 5' end, a unique restriction site near the 3' end, followed by the TAAT stop sequence and a Hind III restriction enzyme site at the very 3' end. The first four oligos, comprising the first 84 amino acids of the carboxy terminal 219 amino acids of human Type I ( $\alpha$ 1) collagen with optimized *E. coli* codon usage, are given in Figure 81 (SEQ. ID. NOS. 47-50).

[0232] Oligos N4-1 (SEQ. ID. NO. 47) and N4-2 (SEQ. ID. NO. 48) (1 μg each) were annealed in 20 μL of T7 DNA polymerase buffer (40mM Tris-HCI (pH 8.0), 5mM MgCl<sub>2</sub>, 5mM dithiothreitol, 50mM NaCl, 0.05 mg/mL bovine serum albumin) by heating at 90°C for 5 minutes followed by slow cooling to room temperature. After brief centrifugation at 14,000 rpm, 10 units of T7 DNA polymase and 2 µL of a solution of all four dNTPs (dATP, dGTP, dCTP, dTTP, 2.5mM each) were added to the annealed oligos. Extension reactions were incubated at 37°C for 30 minutes and then heated at 70°C for 10 minutes. After cooling to room temperature, Hind III buffer (5 µL of 10 x concentration), 20 µL of H<sub>2</sub>O, and 10 units of Hind III restriction enzyme were added and the tubes incubated at 37°C for 10 hours. Hind III buffer (2 μL of 10x concentration), 13.5 μL of 0.5M Tris HCl (pH 7.5), 1.8 μL of 1% Triton X100, 5.6 μL of H20, and 20 U of EcoR I were added to each tube and incubation continued for 2 hours at 37°C. Digests were extracted once with an equal volume of phenol, once with phenol/chloroform/isoamyl alcohol, and once with chloroform/isoamyl alcohol. After ethanol precipitation, the pellet was resuspended in 10 µL of TE buffer (10 mM Tris HCI (pH 8.0), 1 mM EDTA). Resuspended pellet 4 μL of was ligated overnight at 16°C with agarose gel-purified EcoRI/Hind III digested pBSKS+ vector (1 μg) using T4 DNA ligase (100 units). One half of the transformation mixture was transformed by heat shock into DH5\(\pi\) cells and 100 μL of the 1.0 mL transformation mixture was plated on Luria Broth (LB) agar plates containing 70 μg/mL ampicillin. Plates were incubated overnight at 37°C. Ampicillin resistant colonies (6-12) were picked and grown overnight in LB media containing 70μg/mL ampicillin. Plasmid DNA was isolated from each culture by Wizard Minipreps (Promega Corporation, Madison WI) and screened for the presence of the approximately 120 base pair insert by digestion with EcoRI and Hind III and running the digestion products on agarose electrophoresis gels. Clones with inserts were confirmed by standard dideoxy termination DNA sequencing. The correct clone was named pBSN4-1.

[0233] Oligos N4-3 (SEQ. ID. NO. 49) and N4-4 (SEQ. ID. NO. 50) (Figure 81) were synthesized, purified, annealed, extended, and cloned into pBSKS<sup>+</sup> following exactly the same procedure given above for oligos N4-1 and N4-2. The resulting plasmid was named pBSN4-2A. To clone together the sections of the collagen gene from pBSN4-1 and pBSN4-2A, plasmid pBSN4-1 (1μg) was digested for 2 hours at 37°C with Apa L1 and Hind III. The digested vector was purified by agarose gel electrophoresis. Plasmid pBSN4-2A (3 μg) was digested for 2 hours at 37°C with Apa L1 and Hind III and the insert purified by agarose gel electrophoresis. Apa L1/Hind III-digested pBSN4-1 was ligated with this insert overnight at 16°C with T4 DNA ligase. One half of the ligation mixture was transformed into DH5α cells and 1/10 of the transformation mixture was plated on LB agar plates containing 70 μg/mL ampicillin. After overnight incubation at 37°C, ampicillin-resistant clones were picked and screened for the presence of insert DNA as described above. Clones were confirmed by dideoxy termination sequencing. The correct clone was named pBSN4-2.

[0234] In a similar manner, the remainder of the gene for the carboxy terminal 219 amino acids of human Type I ( $\alpha$ 1) collagen with optimized *E. coli* codon usage was constructed such that the final DNA sequence is that given in Figure 71 (SEQ. ID. NO. 43).

[0235] It will be understood that various modifications may be made to the embodiments disclosed herein. For example, it is contemplated that any protein produced by prokaryotes and eukaryotes can be made to incorporate one or more amino acid analogs in accordance with the present disclosure. Therefore, the above description should not be construed as limiting, but merely as exemplifications of preferred embodiments. Those skilled in art will envision other modifications within the scope and spirit of the claims appended hereto.

# Annex to the description

[0236]

5

SEQUENCE LISTING

10	(1) GENERAL INFORMATION:
a j	(i) APPLICANT: GRUSKIN, ELLIOT A.
15	BUECHTER, DOUGLAS
	BROKAW, JANE
	ZHANG, GUANGHUI
20	PAOLELLA, DAVID
25	(ii) TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
23	(iii) NUMBER OF SEQUENCES: 50
30	(iv) CORRESPONDENCE ADDRESS:
30	(A) ADDRESSEE: DILWORTH & BARRESE
	(B) STREET: 333 EARLE OVINGTON BOULEVARD
4.	(C) CITY: UNIONDALE
35	(D) STATE: NY
	(E) COUNTRY: U.S.A.
. •	(F) ZIP: 11553
40	
	(v) COMPUTER READABLE FORM:
	(A) MEDIUM TYPE: Floppy disk
45	(B) COMPUTER: IBM PC compatible
	(C) OPERATING SYSTEM: PC-DOS/MS-DOS
	(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
50	(vi) CURRENT APPLICATION DATA:
	(A) APPLICATION NUMBER:
	(B) FILING DATE:
55	(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

	(A) NAME: STEEN, JEFFREY S	
. 5		
Ç.	(ix) TELECOMMUNICATION INFORMATION:	
	(A) TELEPHONE: (516) 228-8484	
10	(B) TELEFAX: (516) 228-8516	
. 15	(2) INFORMATION FOR SEQ ID NO:1:	
	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 3170 base pairs	
.:	(B) TYPE: nucleic acid	•
•	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
25		
	(ii) MOLECULE TYPE: cDNA	
	( I) anomyga paggarattay gro ta yo i	
30 -	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:	
<i>ħ</i> ∗	CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC TGGCCCCATG	60
	CAGCIGICII AIGGCIAIGA IGAGAANICA ACCGGAGGAA IIICCOICCO IGCCCCAIC	
35	GGTCCCTCTG GTCCTCGTGG TCTCCCTGGC CCCCCTGGTG CACCTGGTCC CCAAGGCTTC	120
•	G32CC2.C.O. G.G0.CC2.CC	
	CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT	180
40		
	CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAA CCTGGAAAACC TGGTCGTCCT	240
45	GGTGAGCGTG GGCCTCCTGG GCCTCAGGGT GCTCGAGGAT TGCCCGGAAC AGCTGGCCTC	300
:		
	CCTGGAATGA AGGGACACAG AGGTTTCAGT GGTTTGGATG GTGCCAAGGG AGATGCTGGT	360
50		
÷.	CCTGCTGGTC CTAAGGGTGA GCCTGGCAGC CCTGGTGAAA ATGGAGCTCC TGGTCAGATG	420
-		

GGCCCCCGTG	GCCTGCCTGG	TGAGAGAGGT	CGCCCTGGAG	CCCCTGGCCC	TGCTGGTGCT	480
CGTGGAAATG	ATGGTGCTAC	TGGTGCTGCC	GGCCCCCTG	GTCCCACCGG	CCCCGCTGGT	540
CCTCCTGGCT	TCCCTGGTGC	TGTTGGTGCT	AAGGGTGAAG	CTGGTCCCCA	AGGGCCCCGA	600
GGCTCTGAAG	GTCCCCAGGG	TGTGCGTGGT	GAGCCTGGCC	CCCCTGGCCC	TGCTGGTGCT	660
GCTGGCCCTG	CTGGAAACCC	TGGTGCTGAT	GGACAGCCTG	GTGCTAAAGG	TGCCAATGGT	720
GCTCCTGGTA	TTGCTGGTGC	TCCTGGCTTC	CCTGGTGCCC	GAGGCCCCTC	TGGACCCCAG	780
GGCCCCGGCG	GCCCTCCTGG	TCCCAAGGGT	AACAGCGGTG	AACCTGGTGC	TCCTGGCAGC	840
AAAGGAGACA	CTGGTGCTAA	GGGAGAGCCT	GGCCCTGTTG	GTGTTCAAGG	ACCCCCTGGC	900
CCTGCTGGAG	AGGAAGGAAA	GCGAGGAGCT	CGAGGTGAAC	CCGGACCCAC	TGGCCTGCCC	960
GGACCCCCTG	GCGAGCGTGG	TGGACCTGGT	AGCCGTGGTT	TCCCTGGCGC	AGATGGTGTT	1020
GCTGGTCCCA	AGGGTCCCGC	TGGTGAACGT	GGTTCTCCTG	GCCCCGCTGG	CCCCAAAGGA	1080
TCTCCTGGTG	AAGCTGGTCG	TCCCGGTGAA	GCTGGTCTGC	CTGGTGCCAA	GGGTCTGACT	1140
GGAAGCCCTG	GCAGCCCTGG	TCCTGATGGC	AAAACTGGCC	CCCCTGGTCC	CGCCGGTCAA	1200
GATGGTCGCC	CCGGACCCCC	AGGCCCACCT	GGTGCCCGTG	GTCAGGCTGG	TGTGATGGGA	1260
TTCCCTGGAC	CTAAAGGTGC	TGCTGGAGAG	CCCGGCAAGG	CTGGAGAGCG	AGGTGTTCCC	1320
GGACCCCCTG	GCGCTGTCGG	TCCTGCTGGC	AAAGATGGAG	AGGCTGGAGC	TCAGGGACCC	1380
CCTGGCCCTG	CTGGTCCCGC	TGGCGAGAGA	GGTGAACAAG	GCCCTGCTGG	CTCCCCCGGA	1440

	TTCCAGGGTC	TCCCTGGTCC	TGCTGGTCCT	CCAGGTGAAG	CAGGCAAACC	TGGTGAACAG	1500
5	GGTGTTCCTG	GAGACCTTGG	CGCCCCTGGC	CCCTCTGGAG	CAAGAGGCGA	GAGAGGTTTC	1560
40	CCTGGCGAGC	GTGGTGTGCA	AGGTCCCCCT	GGTCCTGCTG	GACCCCGAGG	GGCCAACGGT	1620
10	GCTCCCGGCA	ACGATGGTGC	TAAGGGTGAT	GCTGGTGCCC	CTGGAGCTCC	CGGTAGCCAG	1680
15	GGCGCCCCTG	GCCTTCAGGG	AATGCCTGGT	GAACGTGGTG	CAGCTGGTCT	TCCAGGGCCT	1740
5	AAGGGTGACA	GAGGTGATGC	TGGTCCCAAA	GGTGCTGATG	GCTCTCCTGG	CAAAGATGGC	1800
20	GTCCGTGGTC	TGACCGGCCC	CATTGGTCCT	CCTGGCCCTG	CTGGTGCCCC	TGGTGACAAG	1860
25	GGTGAAAGTG	GTCCCAGCGG	CCCTGCTGGT	CCCACTGGAG	CTCGTGGTGC	CCCCGGAGAC	1920
	CGTGGTGAGC	CTGGTCCCCC	CGGCCCTGCT	GGCTTTGCTG	GCCCCCTGG	TGCTGACGGC	1980
30	CAACCTGGTG	CTAAAGGCGA	ACCTGGTGAT	GCTGGTGCCA	AAGGCGATGC	TGGTCCCCCT	2040
	GGGCCTGCCG	GACCCGCTGG	ACCCCCTGGC	CCCATTGGTA	ATGTTGGTGC	TCCTGGAGCC	2100
35	AAAGGTGCTC	GGGCAGCGCT	GGTCCCCCTG	GTGCTACTGG	TTTCCCTGGT	GCTGCTGGCC	2160
: 40	GAGTCGGTCC	TCCTGGCCCC	TCTGGAAATG	CTGGACCCCC	TGGCCCTCCT	GGTCCTGCTG	2220
	GCAAAGAAGG	CGGCAAAGGT	CCCCGTGGTG	AGACTGGCCC	TGCTGGACGT	CCTGGTGAAG	2280
45 <sup>:</sup>	TTGGTCCCCC	TGGTCCCCCT	GGCCCTGCTG	GCGAGAAAGG	ATCCCCTGGT	GCTGATGGTC	2340
	CTGCTGGTGC	TCCTGGTACT	CCCGGGCCTC	AAGGTATTGC	TGGACAGCGT	GGTGTGGTCG	2400
50	GCCTGCCTGG	TCAGAGAGGA	GAGAGAGGCT	TCCCTGGTCT	TCCTGGCCCC	TCTGGTGAAC	2460

CTGGCAAACA	AGGTCCCTCT	GGAGCAAGTG	GTGAACGTGG	TCCCCCGGT	CCCATGGGCC	2520
CCCCTGGATT	GGCTGGACCC	CCTGGTGAAT	CTGGACGTĢA	GGGGGCTCCT	GCTGCCGAAG	2580
GTTCCCCTGG	ACGAGACGGT	TCTCCTGGCG	CCAAGGGTGA	CCGTGGTGAG	ACCGGCCCCG	2640
CTGGACCCCC	TGGTGCTCCT	GGTGCTCCTG	GTGCCCCTGG	CCCCGTTGGC	CCTGCTGGCA	2700
AGAGTGGTGA	TCGTGGTGAG	ACTGGTCCTG	CTGGTCCCGC	CGGTCCCGTC	GGCCCCGCTG	2760
GCGCCCGTGG	CCCCGCCGGA	CCCCAAGGCC	CCCGTGGTGA	CAAGGGTGAG	ACAGGCGAAC	2820
AGGGCGACAG	AGGCATAAAG	GGTCACCGTG	GCTTCTCTGG	CCTCCAGGGT	CCCCTGGCC	2880
CTCCTGGCTC	TCCTGGTGAA	CAAGGTCCCT	CTGGAGCCTC	TGGTCCTGCT	GGTCCCCGAG	2940
GTCCCCCTGG	CTCTGCTGGT	GCTCCTGGCA	AAGATGGACT	CAACGGTCTC	CCTGGCCCCA	3000
TTGGGCCCCC	TGGTCCTCGC	GGTCGCACTG	GTGATGCTGG	TCCTGTTGGT	ccccccccc	3060
CTCCTGGACC	TCCTGGTCCC	CCTGGTCCTC	CCAGCGCTGG	TTTCGACTTC	AGCTTCCTCC	3120
CCCAGCCACC	TCAAGAGAAG	GCTCACGATG	GTGGCCGCTA	CTACCGGGCT		3170

#### (2) INFORMATION FOR SEQ ID NO:2:

5

10

15

20

25

30

35

40

45

50

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 240 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:	
.; 5	CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC TGGCCCCATG	60
51 10	GGTCCCTCTG GTCCTCGTGG TCTCCCTGGC CCCCCTGGTG CACCTGGTCC CCAAGGCTTC	120
	CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT	180
15	CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT	240
20	(2) INFORMATION FOR SEQ ID NO:3:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 100 base pairs	
25	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
30	(ii) MOLECULE TYPE: cDNA	
హా <b>35</b> ల	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:	
	GGATCCATGG GGCTCGCTGG CCCACCGGGC GAACCGGGTC CGCCAGGCCC GAAAGGTCCG	60
40	CGTGGCGATA GCGGGCTCCC GGGCGATTCC TAATGGATCC	100

; 

	(2)	INFO	RMATI	ON FOR	SEQ ID	NO:4:									
5		(4)	SROIT	ENCE CH	ARACTE	etsttcs	•								
		(2)		LENGTH				•							
				TYPE:			148								
10				STRAND											
				TOPOLO			<b>C</b>								
		•	(D)	TOPOLO	or, win	HOWIT									
15		(ii)	MOLE	CULE TY	PE: per	otide									
٠		(xi)	SEQU	ENCE DE	SCRIPTI	ON: SE	Q ID NO	:4:							
20		<b>01</b>	7	NI - 01	Des De		<b>01</b> Desc	<b>a</b> 1	D	D	<b>01</b>	<b>D</b>	<b>-</b>		
,			Leu .	Ala Giy		o Gly	Glu Pro		Pro	Pro	GIĀ	Pro	•	GIÀ	•
		1			5			10					15		
25		Pro	Arg	Gly Asp	Ser										
				20											
30	(2)	INFO	RMATI	ON FOR	SEQ ID	NO:5:									
		(i)	SEQU	ENCE CH	ARACTE	RISTICS	:								
35			(A)	LENGTH	: 330 h	ase pa	irs								
			(B)	TYPE:	nucleio	acid:									
			(C)	STRAND	EDNESS:	singl	e								
40 <sup>-</sup>			(D)	TOPOLO	GY: lir	ıear									
		(ii)	MOLE	CULE TY	PE: cD1	IA									
45		(xi)	SEQU	ENCE DE	SCRIPTI	ION: SE	Q ID NO	: 5 :							
50	CAG	CGGGC	CA GG	AAGAAGA	A TAAGA	VACTGC	CGGCGCCI	ACT (	CGCTC	TATG	T GG	ACTI	CAGO	:	60
	GAT	GTGGG	CT GG	AATGACT	G GATTO	<b>₹</b> TGGCC	CCACCAGO	GCT A	ACCAG	GCCT	т ст	ACTG	CCAT		120

,	GGGGACTG	cc co	CTTTC	CACI	GGG	CTGA	CCAC	CTC	AACT(	CAA	CCAA	CCAT	GC C	ATTG	TGCA	G	180
5	ACCCTGGT	CA A	rtcto	etcaj	TT	CAG'	ratc	ccc	AAAGO	CÇT	GTTG	TGTG(	cc c	ACTG.	AACT	3	240
10	AGTGCCAT	CT C	CATGO	CTGTA	cci	rgga:	rgag	TAT	GATAJ	AGG	TGGT.	ACTG	AA A	AATT.	ATCA	3	300
	GAGATGGT	AG T	AGAGO	GATO	TGO	GTG	CCGC										330
15	(2) INFO	RMAT:	ION I	FOR S	EQ :	ID N	0:6:										
	(i)	SEQU	JENCI	E CH	RACI	reri:	STICS	3:									
20		(A)	LEN	GTH :	116	59 ar	nino	acio	ds								
•		(B)	TYI	PE: a	mino	o ac	id										
•		(C)	STE	RANDE	DNE	5S: £	singl	le									
25		(D)	TOP	POLOG	¥: ι	ınkno	own										
	(ii)	MOLI	ECULE	TYI	E: p	ept:	ide										
ř.																	
30	(xi)	SEQ	JENCI	E DES	CRI	PTIO	N: SI	EQ II	ОИ С	:6:							
q.	Gln	Leu	Ser	Tyr	Gly	Tyr	Asp	Glu	Lys	Ser	Thr	Gly	Gly	Ile	Ser	Val	
35	1				5					10					15		
	Pro	Gly	Pro	Met	Gly	Pro	Ser	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Pro	Pro	
40				20					25					30			
	Gly	Ala	Pro	Gly	Pro	Gln	Gly	Phe	Gln	Gly	Pro	Pro	Gly	Glu	Pro	Gly	
45			35					40					45				
••	<b>~</b> 3		~1				_										
i e	GIU	Pro	GIA	Ala	ser	GIÀ		Met	Gly	Pro	Arg		Pro	Pro	Gly	Pro	
50		50					55					60					
50	Pro	Gly	T.vs	Asn	Glv	Aen	Acn	GI v	G3 v	בות	Glv-	Lare	D=-	glv.	Ara	Dro	
	65	<b>-</b> -y	<b>-</b> 10		J-7	70	nop	GLY	GIU	via	75	פעה	210	GIY	ALY	80	
•	0.5					. •										5.0	

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu  100 105 110  Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro		Gly	Glu	Arg	Gly	Pro	Pro	Gly	Pro	Gln	Gly	Ala	Arg	Gly	Leu	Pro	Gly
Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu 100 105 105 105 105 100 110 110 110 110						85					90					95	
100 105 110 110 110 110 110 110 110 110																	
100		Thr	Ala	Gly	Leu	Pro	Gly	Met	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly	Leu
Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro 115					100					105					110		
115	10																
Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly 130 135 140  Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala 145 150 160  Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr 165 167 170 175  Gly Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Gly Ala Val Gly Ala Lys Gly 180 185 190  Arg Gly Glu Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val 195 200 205  Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Ala Ala Gly Pro Ala 210 215 220  Ala Pro Gly Ile Ala Gly Ala Pro Gly Pro Gly Pro Gly Ala Lys Gly Ala Asn Gly 225 230 240  Ala Pro Gly Ile Ala Gly Ala Pro Gly Pro Gly Pro Gly Pro Gly Ala Arg Gly Pro		Asp	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Glu	Pro
20  Leu Pro Gly Glu Arg Gly Ala Pro Gly Ala Pro Gly Gln Met Gly Pro Arg Gly Ala 155  Arg Gly Ara Asp Asp Asp Arg Gly Pro Gly Ala Ala Gly Pro Gly Pro Thr 165  Glu Ala Gly Pro Gln Gly Pro Pro Gly Pro Gly Ala Val Gly Ala Lys Gly 180  Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val 195  Gly Arg Gly Glu Pro Gly Pro Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val 195  Gly Arg Gly Glu Pro Gly Pro Gly Pro Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val 195  Gly Arg Gly Glu Pro Gly Pro Gly Pro Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val 195  Arg Gly Arg Gly Rro Gly Ala Arg Gly Pro Gly Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala 225  Ala Pro Gly Ile Ala Gly Ala Pro Gly Pro Gly Pro Gly Ala Lys Gly Ala Arg Gly Pro Ala 235	h.			115					120					125			
130	15																
Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala 145		Gly	Ser	Pro	Gly	Glu	Asn	Gly	Ala	Pro	Gly	Gln	Met	Gly	Pro	Arg	Gly
Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala 145			130					135					140				
145	20																
25  Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr 165  Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly 180  Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val 195  Arg Gly Glu Pro Gly Pro Gly Pro Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val 200  Arg Gly Glu Pro Gly Pro Gly Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala 210  Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly 225  Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro		Leu	Pro	Gly	Glu	Arg	Gly	Arg	Pro	Gly	Ala	Pro	Gly	Pro	Ala	Gly	Ala
Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr  165		145					150					155					160
30 Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly 180  Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val 195  Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val 200  Arg Gly Glu Pro Gly Pro Gly Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala 215  Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly 225  Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro 50	25																
Gly Pro Ala Gly Pro Pro Gly Pro Gly Ala Val Gly Ala Lys Gly 180		Arg	Gly	Asn	Asp	Gly	Ala	Thr	Gly	Ala	Ala	Gly	Pro	Pro	Gly	Pro	Thr
Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly  180						165					170					175	
Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly  180	30													_			
Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val 195 200 205  Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala 210 215 220  Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly 225 230 235 240  Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro	-	Gly	Pro	Ala	_	Pro	Pro	Gly	Phe		Gly	Ala	Val	Gly		Lys	Gly
35 Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val 195					180					185					190		
200 205 205 206 207 207 207 208 209 209 209 209 209 209 209 209 209 209					_	<b>-</b>		_	_		_				<b></b>		
Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala 210 215 220  Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly 225 230 235 240  Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro	35	Glu	Ala		Pro	Gin	Gly	Pro		GIÀ	Ser	GIU	GIĀ		GID	GIA	Val
Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala 210 215 220  Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly 225 230 235 240  Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro	•			195					200					205			
210 215 220  Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly 225 230 235 240  Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro	ris.	•	<b>61.</b> 4	<b>~1</b>	Dwa	~1·	Dvo	Dwa	~1··	Dro	21-	G1 v	21.0	מות	Glv.	Dro	21-
Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly 225 230 235 240  Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro	40	Arg		GIU	PIO	GIY	PIO		GIY	PLO	MIG	GIY		MIG	GIY	PLO	ALG
225 230 235 240  Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro			210					213					220				
225 230 235 240  Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro		G] v	Δen	Pro	Glv	Ala	Asp	Glv	Gln	Pro	G1v	Ala	Lvg	Glv	Ala	Asn	Gl v
Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro	45	_	79**		<b>-</b>		_	UL,	·		U.,		-,,	OL,	,,,,,		
50		223					250										
50		Ala	Pro	Glv	Ile	Ala	Glv	Ala	Pro	Glv	Phe	Pro	Glv	Ala	Arg	Gly	Pro
	50			- 4			-3			4					•		

·	Ser	Gly	Pro	Gln	Gly	Pro	Gly	Gly	Pro	Pro	Gly	Pro	Lys	Gly	Asn	Ser
5				260					265					270		
	Gly	Glu	Pro	Gly	Ala	Pro	Gly	Ser	Lys	Gly	Asp	Thr	Gly	Ala	Lys	Gly
			275					280					285			
10																
,	Glu	Pro	Gly	Pro	Val	Gly	Val	Gln	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Glu
:		290					295					300				
15																
	Glu	Gly	Lys	Arg	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Pro	Thr	Gly	Leu	Pro
	305					310					315					320
20																
	Gly	Pro	Pro	Gly	Glu	Arg	Gly	Gly	Pro	Gly	Ser	Arg	Gly	Phe	Pro	Gly
					325					330					335	
25																
•	Ala	Asp	Gly		Ala	Gly	Pro	Lys	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Ser
				340					345					350		
30		_						_			_	_	_			
	Pro	Gly		Ala	Gly	Pro	Lys	_	Ser	Pro	Gly	Glu		Gly	Arg	Pro
·			355					360					365			
35	<b>61</b>	<b>01</b>	21-	<b>61</b>	<b>7</b>	D	<b>a</b> 1		•	<b>~1</b>	•	m\	01		_	
	GIY		Ala	GIÀ	Leu	Pro		Ala	гÀв	GIĀ	Leu		GIÀ	Ser	Pro	GIÀ
		370					375					380				
w 1	Car	Dro	alv	Pro	yen.	Gl v	Lva	The s	<i>0</i> 1	Dro	Dro	01.	Dro	Ala	<b>01</b>	<b>~1</b> -
40	385	PLO	GIŞ	PLO	Map	390	пÃя	IIII	GIÀ	PIQ	395	GIY	PIO	ATS	GTÅ	
	303					390					393					400
	Nen.	Gly	Ara	Pro	Glv	Pro	Dro	Gl v	Pro	Dro	Gl v		7.~~	Gly	C15	21-
45	App	Gry	AT 9	710	405	FIO	FLO	GIY	FIO	410	GIŞ	ATG	wig	GIY	415	мта
					403				•	410					413	
	Glv	Val	Met	G] v	Phe	Pro	Glv	Pro	Lve	Glv	בומ	בומ	Glv	Glu	Pro	G] v
50	7			420		0			425	U1 y	,a	nia	GL Y	430		JIY
														-70		

	Lys	Ala	Gly	Glu	Arg	Gly	Val	Pro	Gly	Pro	Pro	Gly	Ala	Val	Gly	Pro
5			435					440					445			
	Ala	Gly	Lys	Asp	Gly	Glu	Ala	Gly	Ala	Gln	Gly	Pro	Pro	Gly	Pro	Ala
		450	•	-	-		455	•	•		_	460		_		
10																
	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Glu	Gln	Gly	Pro	Ala	Gly	Ser	Pro	Gly
	465					470					475					480
15	_,	~3	<b>a1</b>	•	D	<b>0</b> 7	D	.1-	<b>01</b>	Dua	D==	<b>a</b> 1	<b>71</b>	210	<b>01.</b> .	•
	Phe	Gln	GTÅ	Leu	485	GIY	Pro	AIA	-	490	PIO	GIY	Giu	AIA	495	гув
20	Pro	Gly	Glu	Gln	Gly	Val	Pro	Gly	Asp	Leu	Gly	Ala	Pro	Gly	Pro	Ser
				500					505					510		
25																
	Gly	Ala	-	Gly	Glu	Arg	Gly		Pro	Gly	Glu	Arg		Val	Gln	Gly
			515					520					525			
30	Pro	Pro	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Ala	Asn	Gly	Ala	Pro	Gly	Asn
		530					535					540				
35	Asp	Gly	Ala	Lys	Gly		Ala	Gly	Ala	Pro		Ala	Pro	Gly	Ser	
	545					550					555					560
	Glv	Ala	Pro	Glv	Leu	Gln	Glv	Met	Pro	Glv	Glu	Arq	Glv	Ala	Ala	Glv
40	OLY			0-7	565		,			570		<b>J</b>	2		575	,
45	Leu	Pro	Gly	Pro	Lys	Gly	Asp	Arg	Gly	Asp	Ala	Gly	Pro	Lys	Gly	Ala
45				580					585					590		
	<b>3</b>	03	Corr	n	<b>~1.</b> .	T	N	G1	17-7	X	C1	T 011	ጥኤ⊶	Cl.	D~~	T10
50	Asp	Gly	595		GIĀ	гув	Asp	600	val	Arg	GIĀ	red	605	GIÅ	PLO	TIG
= -								500								

	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Pro	Gly	Asp	Lys	Gly	Glu	Ser	Gly
5		610					615					620				
	Pro	Ser	Gly	Pro	Ala	-	Pro	Thr	Gly	Ala	Arg	Gly	Ala	Pro	Gly	qaA
10	625					630					635					640
•	Arg	Gly	Glu	Pro	_	Pro	Pro	Gly	Pro		Gly	Phe	Ala	Gly		Pro
15					645					650					655	
	Gly	Ala	Asp	Gly 660	Gln	Pro	Gly	Ala	Lys 665	Gly	Glu	Pro	Gly	<b>A</b> sp 670	Ala	Gly
20				000					003					0.0		
	Ala	Lys	Gly 675	Asp	Ala	Gly	Pro	Pro 680	Gly	Pro	Ala	Gly	Pro 685	Ala	Gly	Pro
25	Pro	Gly	Pro	Ile	Glv	Asn	Va1	Glv	Ala	Pro	Glv	Ala	Lvs	Glv	Ala	Arg
•	220	690			,		695	,				700	-,-	4		<b>J</b>
30	Gly	Ser	Ala	Gly	Pro	Pro	Gly	Ala	Thr	Gly	Phe	Pro	Gly	Ala	Ala	Gly
	705					710					715					720
35'	Arg	Val	Gly	Pro	Pro	Gly	Pro	Ser	Gly	Asn	Ala	Gly	Pro	Pro	Gly	Pro
1					725					730					735	
40	Pro	Gly	Pro	Ala 740	Gly	Lys	Glu	Gly	Gly 745	Lys	Gly	Pro	Arg	Gly 750	Glu	Thr
45	Gly	Pro			Arg	Pro	Gly			Gly	Pro	Pro		Pro	Pro	Gly
			755					760					765			
50	Pro	Ala 770	_	Glu	Lys	Gly	Ser 775		Gly	Ala	Asp	Gly 780		Ala	Gly	Ala

	Pro	Gly	Thr	Pro	Gly	Pro	Gln	Gly	Ile	Ala	Gly	Gln	Arg	Gly	Val	Val
5	785					790					795					800
	Gly	Leu	Pro	Gly	Gln 805	Arg	Gly	Glu	Arg	Gly 810	Phe	Pro	Gly	Leu	Pro 815	Gly
10															023	
	Pro	Ser	Gly		Pro	Gly	Lys	Gln		Pro	Ser	Gly	Ala		Gly	Glu
.15				820					825					830		
	Arg	Gly	Pro	Pro	Gly	Pro	Met	Gly	Pro	Pro	Gly	Leu	Ala	Gly	Pro	Pro
			835					840					845			
20	Gly	Glu	Ser	Gly	Arg	Glu	Gly	Ala	Pro	Ala	Ala	Glu	Gly	Ser	Pro	Gly
		850					855					860				
25	Arg	Asp	Gly	Ser	Pro	Gly	Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro
	865					870					875					880
30	Ala	Glv	Pro	Pro	Gly	Ala	Xaa	Glv	Ala	Xaa	Gly	Ala	Pro	Gly	Pro	Val
					885					890	•			•	895	
35	<b>~</b> 1	Dwa	21-	C1	Lys	Com	a1.,	n an	N ====	Cl.r	<b>~1.</b>	The	<b>Cl</b> 11	Dwa	21-	<b>61</b>
33	GIY	PIO	Ala	900	тЛя	ser	GIY	Авр	905	GIÀ	GIU	Int	GIY	910	ATA	GIÀ
40	Pro	Ala	Gly 915	Pro	Val	Gly	Pro	Ala 920	Gly	Ala	Arg	Gly	Pro 925	Ala	Gly	Pro
			713					720					723			
45	Gln	Gly	Pro	Arg	Gly	Asp	Lys	Gly	Glu	Thr	Gly	Glu	Gln	Gly	Asp	Arg
		930					935	-				940				
	Gly	Ile	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly	Leu	Gln	Gly	Pro	Pro	Gly
50	945					950					955					960

5	Pro	Pro	Gly	Ser	Pro 965	Gly	Glu	Gln	Gly	Pro 970	Ser	Gly	Ala	Ser	Gly 975	Pro
10	Ala	Gly	Pro	Arg 980	Gly	Pro	Pro	Gly	Ser 985	Ala	Gly	Ala	Pro	Gly 990	Lys	Asp
15	Gly	Leu	Asn 995	Gly	Leu	Pro	Gly	Pro 1000		Gly	Pro	Pro	Gly 1005		Arg	Gly
20	Arg	Thr		Asp	Ala	Gly	Pro 1015		Gly	Pro	Pro	Gly 1020		Pro	Gly	Pro
25	Pro 1025		Pro	Pro	Gly	Pro 1030		Ser	Ala	Gly	Phe 1035		Phe	Ser	Phe	Leu 1040
i.	Pro	Gln	Pro	Pro	Gln 1045		Lys	Ala	His	Asp 1050	_	Gly	Arg	Tyr	Tyr 1055	-
00																
<b>30</b>	Ala	Arg	Ser	Gln 1060		Ala	Arg	Lys	<b>Lys</b> 1065		Lys	Asn	Cys	Arg 1070	Arg )	His
				1060 Val	)			_	1065	5			-	1070	_	
<b>6</b> .	Ser	Leu	Tyr 1075 Pro	1060 Val	Asp	Phe	Ser	Asp 1080	1065 Val	Gly	Trp	Asn	Asp 1085	1070 Trp	)	Val
<b>35</b>	Ser Ala	Leu Pro 1090	Tyr 1075 Pro	Val	Азр	Phe Gln	Ser Ala 1095 Asn	Asp 1080 Phe	Val	Gly	Trp	Asn Gly 1100	Asp 1085 Asp	Trp	Ile	Val Phe

	Thr	Glu	Leu	Ser	Ala	Ile	Ser	Met	Leu	Tyr	Leu	Asp	Glu	Tyr	Asp	Lys	
5				1140	)				1149	5				1150	0		
,, 10	Val	Val	Leu 1155		Asn	Tyr	Gln	Glu 1160		Val	Val	Glu	Gly		Gly	Сув	
	Arg																
15	(2) INFO	RMAT]	ON I	FOR S	SEQ 1	ID NO	0:7:										
20	(i)		LEN	IGTH:	353	31 ba	ase p		3								
25	·	(C)	STE	PE: I	DNES	SS: s	sing!	le									
30	(ii)	MOLI	ECULE	3 TYE	E: c	DNA											
	(xi)	SEQU	JENCE	DES	CRIE	PTION	1: SI	EQ II	O NO:	7:							
3 <b>5</b> 1.	GGGAAGGA'	TT TO	CATT	rtccc	: AGC	CTGTO	CTTA	TGG	CTATO	AT (	gagai	\ATCA	A C	GGAG	GAAT	ŗ ·	60
<u>;</u> ,	TTCCGTGC	CT GO	CCCC	CATGO	GTO	CCTC	CTGG	TCC	rcgte	GT (	CTCC	CTGGC	:c c	CCTC	GTGC	2	120
40	ACCTGGTC	CC C2	\AGG(	CTTCC	: AAG	GTC	ccc	TGGT	rgago	CT (	GCG1	AGCCT	G G	AGCTT	CAGG	<del>}</del>	180
45	TCCCATGG	GT C	CCGA	AGGTC	ccc	CAGG	STCC	ccc	rggaa	AG I	AATG	BAGAT	G A	rgggg	BAAGO	2	240
	TGGAAAAC	CT GO	STCGI	CCTG	GTO	BAGC	STGG	GCC	CCTG	GG (	CCTC	AGGGT	G CI	CGAG	GATI	r	300
50	GCCCGGAA	CA GO	CTGGC	CTCC	сто	GAAT	rgaa	GGG7	ACACA	GA (	GGTT	CAGI	'G G7	TTTG	SATGO	3	360
1	TGCCAAGG	GA G	ATGCI	rggto	CTC	CTG	STCC	TAAC	GGTG	ag (	CCTG	CAGO	C C	rggte	BAAA	١.	420

;

·;; 

480	GCCCTGGAGC	GAGAGAGGTC	CCTGCCTGGT	GCCCCCGTGG	GGTCAGATGG	TGGAGCTCCT	
540	GGCCCCCTGG	GGTGCTGCCG	TGGTGCTAÇT	GTGGAAATGA	GCTGGTGCTC	CCCTGGCCCT	
600	AGGGTGAAGC	GTTGGTGCTA	CCCTGGTGCT	CTCCTGGCTT	CCCGCTGGTC	TCCCACCGGC	
660	AGCCTGGCCC	GTGCGTGGTG	TCCCCAGGGT	GCTCTGAAGG	GGGCCCCGAG	TGGTCCCCAA	
720	GACAGCCTGG	GGTGCTGATG	TGGAAACCCT	CTGGCCCTGC	GCTGGTGCTG	CCCTGGCCCT	
780	CTGGTGCCCG	CCTGGCTTCC	TGCTGGTGCT	CTCCTGGTAT	GCCAATGGTG	TGCTAAAGGT	
·840	ACAGCGGTGA	CCCAAGGGTA	CCCTCCTGGT	GCCCCGGCGG	GGACCCCAGG	AGGCCCCTCT	
900	GCCCTGTTGG	GGAGAGCCTG	TGGTGCTAAG	AAGGAGACAC	CCTGGCAGCA	ACCTGGTGCT	
960	GAGGTGAACC	CGAGGAGCTC	GGAAGGAAAG	CTGCTGGAGA	CCCCCTGGCC	TGTTCAAGGA	
1020	GCCGTGGTTT	GGACCTGGTA	CGAGCGTGGT	GACCCCCTGG	GGCCTGCCCG	CGGACCCACT	
1080	GTTCTCCTGG	GGTGAACGTG	GGGTCCCGCT	CTGGTCCCAA	GATGGTGTTG	CCCTGGCGCA	
1140	CTGGTCTGCC	CCCGGTGAAG	AGCTGGTCGT	CTCCTGGTGA	CCCAAAGGAT	CCCCGCTGGC	
1200	AAACTGGCCC	CCTGATGGCA	CAGCCCTGGT	GAAGCCCTGG	GGTCTGACTG	TGGTGCCAAG	
1260	GTGCCCGTGG	GGCCCACCTG	CGGACCCCCA	ATGGTCGCCC	GCCGGTCAAG	CCCTGGTCCC	
1320	CCGGCAAGGC	GCTGGAGAGC	TAAAGGTGCT	TCCCTGGACC	GTGATGGGAT	TCAGGCTGGT	
1380	AAGATGGAGA	CCTGCTGGCA	CGCTGTCGGT	GACCCCCTGG	GGTGTTCCCG	TGGAGAGCGA	
1440	GTGAACAAGG	GGCGAGAGAG	TGGTCCCGCT	CTGGCCCTGC	CAGGGACCCC	GGCTGGAGCT	

CCCTGCTGGC	TCCCCCGGAT	TCCAGGGTCT	CCCTGGTCCT	GCTGGTCCTC	CAGGTGAAGC	1500
AGGCAAACCT	GGTGAACAGG	GTGTTCCTGG	AGACCTTGGC	GCCCCTGGCC	CCTCTGGAGC	1560
AAGAGGCGAG	AGAGGTTTCC	CTGGCGAGCG	TGGTGTGCAA	GGTCCCCCTG	GTCCTGCTGG	1620
ACCCCGAGGG	GCCAACGGTG	CTCCCGGCAA	CGATGGTGCT	AAGGGTGATG	CTGGTGCCCC	1680
TGGAGCTCCC	GGTAGCCAGG	GCGCCCTGG	CCTTCAGGGA	ATGCCTGGTG	AACGTGGTGC	1740
AGCTGGTCTT	CCAGGGCCTA	AGGGTGACAG	AGGTGATGCT	GGTCCCAAAG	GTGCTGATGG	1800
CTCTCCTGGC	AAAGATGGCG	TCCGTGGTCT	GACCGGCCCC	ATTGGTCCTC	CTGGCCCTGC	1860
TGGTGCCCCT	GGTGACAAGG	GTGAAAGTGG	TCCCAGCGGC	CCTGCTGGTC	CCACTGGAGC	1920
TCGTGGTGCC	CCCGGAGACC	GTGGTGAGCC	TGGTCCCCC	GGCCCTGCTG	GCTTTGCTGG	1980
CCCCCTGGT	GCTGACGGCC	AACCTGGTGC	TAAAGGCGAA	CCTGGTGATG	CTGGTGCCAA	2040
AGGCGATGGG	TCCCCCTGGG	CCTGCCGGAC	CCGCTGGACC	CCCTGGCCCC	ATTGGTAATG	2100
TTGGTGCTCC	TGGAGCCAAA	GGTGCTCGCG	GCAGCGCTGG	TCCCCCTGGT	GCTACTGGTT	2160
TCCCTGGTGC	TGCTGGCCGA	GTCGGTCCTC	CTGGCCCCTC	TGGAAATGCT	GGACCCCCTG	2220
GCCCTCCTGG	TCCTGCTGGC	AAAGAAGGCG	GCAAAGGTCC	CCGTGGTGAG	ACTGGCCCTG	2280
CTGGACGTCC	TGGTGAAGTT	GGTCCCCCTG	GTCCCCCTGG	CCCTGCTGGC	GAGAAAGGAT	2340
CCCCTGGTGC	TGATGGTCCT	GCTGGTGCTC	CTGGTACTCC	CGGGCCTCAA	GGTATTGCTG	2400
GACAGCGTGG	TGTGGTCGGC	CTGCCTGGTC	AGAGAGGAGA	GAGAGGCTTC	CCTGGTCTTC	2460

	CTGGCCCCTC	TGGTGAACCT	GGCAAACAAG	GTCCCTCTGG	AGCAAGTGGT	GAACGTGGTC	2520
<b>.</b>	CCCCCGGTCC	CATGGGCCCC	CCTGGATTGG	CTGGACCCCC	TGGTGAATCT	GGACGTGAGG	2580
10	GGGCTCCTGC	TGCCGAAGGT	TCCCCTGGAC	GAGACGGTTC	TCCTGGCGCC	AAGGGTGACC	2640
P.,	GTGGTGAGAC	CGGCCCCGCT	GGACCCCCTG	GTGCTCTGGT	GCTCTGGTGC	CCCTGGCCCC	2700
15	GTTGGCCCTG	CTGGCAAGAG	TGGTGATCGT	GGTGAGACTG	GTCCTGCTGG	TCCCGCCGGT	2760
	CCCGTCGGCC	CCGCTGGCGC	CCGTGGCCCC	GCCGGACCCC	AAGGCCCCCG	TGGTGACAAG	2820
20	GGTGAGACAG	GCGAACAGGG	CGACAGAGGC	ATAAAGGGTC	ACCGTGGCTT	CTCTGGCCTC	2880
	CAGGGTCCCC	CTGGCCCTCC	TGGCTCTCCT	GGTGAACAAG	GTCCCTCTGG	AGCCTCTGGT	2940
<b>25</b>	CCTGCTGGTC	CCCGAGGTCC	CCCTGGCTCT	GCTGGTGCTC	CTGGCAAAGA	TGGACTCAAC	3000
30	GGTCTCCCTG	GCCCCATTGG	GCCCCCTGGT	CCTCGCGGTC	GCACTGGTGA	TGCTGGTCCT	3060
3	GTTGGTCCCC	CCGGCCCTCC	TGGACCTCCT	GGTCCCCCTG	GTCCTCCCAG	CGCTGGTTTC	3120
35 <sup>-</sup>	GACTTCAGCT	TCCTCCCCCA	GCCACCTCAA	GAGAAGGCTC	ACGATGGTGG	CCGCTACTAC	3180
.2	CGGGCTAGAT	CCCAGCGGGC	CAGGAAGAAG	AATAAGAACT	GCCGGCGCCA	CTCGCTCTAT	3240
40	GTGGACTTCA	GCGATGTGGG	CTGGAATGAC	TGGATTGTGG	CCCCACCAGG	CTACCAGGCC	3300
45	TTCTACTGCC	ATGGGGACTG	CCCCTTTCCA	CTGGCTGACC	ACCTCAACTC	AACCAACCAT	3360
	GCCATTGTGC	AGACCCTGGT	CAATTCTGTC	AATTCCAGTA	TCCCCAAAGC	CTGTTGTGTG	3420
50	CCCACTGAAC	TGAGTGCCAT	CTCCATGCTG	TACCTGGATG	AGTATGATAA	GGTGGTACTG	3480

5	AAAAATTA	TC AG	GAGA	TGGT	' AGI	AGAG	GGA	TGTG	GGTG	SCC (	CTAP	AAGO	т				3531
	(2) INFO	RMATI	ON F	or s	EQ I	D NO	8:			•							
-10	(i)	SEQU	ENCE LEN						ls								
·15		(C)	TYP STR TOP	ANDE	DNES	SS: s	singl	.e								,	
20	(ii)	MOLE	CULE	TYE	PE: p	epti	ide										
•	(xi)	SEQU	JENCE	DES	CRI	PTION	1: SI	EQ II	NO:	8:		,					•
25	Gln 1	Leu	Ser	Tyr	Gly 5	туг	Asp	Glu	Lys	Ser 10	Thr	Gly	Gly	Ile	Ser 15	Val	
30 <sup>°</sup>	Pro	Gly	Pro	Met 20	Gly	Pro	Ser	Gly	Pro 25	Arg	Gly	Leu	Pro	Gly 30	Pro	Pro	
35	Gly	Ala	Pro 35	Gly	Pro	Gln	Gly	Phe 40	Gln	Gly	Pro	Pro	Gly 45	Glu	Pro	Gly	
40	Glu	Pro	Gly	Ala	Ser	Gly	Pro 55	Met	Gly	Pro	Arg	Gly 60	Pro	Pro	Gly	Pro	
45	Pro	Gly	Lys	Asn	Gly	Asp	Asp	Gly	Glu	Ala	Gly 75	Lys	Pro	Gly	Arg	Pro 80	
50	Gly	<b>Glu</b>	Arg	Gly	Pro 85	Pro	Gly	Pro	Gln	Gly 90	Ala	Arg	Gly	Leu	Pro 95	Gly	
55	Thr	Ala	Gly	Leu 100	Pro	Gly	Met	Lys	Gly 105	His	Arg	Gly	Phe	Ser 110	Gly	Leu	

	Asp	Gly	Ala	Lys	Gly	qaA	Ala	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Glu	Pro
5			115					120					125			
			_						_		-1		~1		_	
	Gly		Pro	GIA	Glu	Asn	_	Ala	Pro	GIY	GIN		GIY	Pro	Arg	Gly
10		130					135					140				
	T.em	Pro	Glv	Glu	Ara	Glv	Ara	Pro	Glv	Ala	Pro	Glv	Pro	Ala	Glv	Ala
	145		<b>5-</b> 7	<b></b>	3	150	5		1	•	155				,	160
15																
	Arg	Gly	Asn	Asp	Gly	Ala	Thr	Gly	Ala	Ala	Gly	Pro	Pro	Gly	Pro	Thr
•					165					170					175	
20																
	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Phe	Pro	Gly	Ala	Val	Gly	Ala	Lys	Gly
				180					185					190		
25							_	_			<b>~1</b>	01		<b>6</b> 3 -	<b>61</b>	
	Glu	Ala	_	Pro	GIn	GIÀ	Pro	Arg	GIÀ	ser	GIU	GIĀ	205	Gln	GIĀ	vaı
			195					200					205			
30	Arg	Glv	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Ala	Gly	Pro	Ala
	5	210			. •		215	•			_	220		-		
35	Gly	Asn	Pro	Gly	Ala	Asp	Gly	Gln	Pro	Gly	Ala	Lys	Gly	Ala	Asn	Gly
	225					230					235					240
•																
40	Ala	Pro	Gly	Ile		Gly	Ala	Pro	Gly			Gly	Ala	Arg		Pro
					245					250					255	
	50*	G) v	Pro	Gln	G] v	Pro	Glv	Glv	Pro	Pro	ദിഴ	Pro	Taya	Gly	Δan	Ser
45	261	GIY	PIO	260	_		Gry	·	265		G.,	210	Lyb	270		001
										•				_, ,		
	Gly	Glu	Pro	Gly	Ala	Pro	Gly	Ser	Lys	Gly	Asp	Thr	Gly	Ala	Lys	Gly
50			275					280					285			

	Glu	Pro	Gly	Pro	Val	Gly	Val	Gln	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Glu
5		290					295					300				
ai	Glu	Gly	Lys	Arg	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Pro	Thr	Gly	Leu	Pro
10	305					310					315					320
	Gly	Pro	Pro	Gly	Glu	Arg	Gly	Gly	Pro	Gly	Ser	Arg	Gly	Phe	Pro	Gly
15					325					330					335	
•	Ala	Asp	Gly		Ala	Gly	Pro	Lys	_	Pro	Ala	Gly	Glu		Gly	Ser
20				340					345					350		
	Pro	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Ser	Pro	Gly	Glu	Ala	Gly	Arg	Pro
25			355					360					365			
•	Gly	Glu	Ala	Gly	Leu	Pro	Gly	Ala	Lys	Gly	Leu	Thr	Gly	Ser	Pro	Gly
30		370				•	375					380				
	Ser	Pro	Gly	Pro	Asp	Gly	Lys	Thr	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Gln
6 <b>35</b> '	385					390					395					400
	Asp	Gly	Arg	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Ala	Arg	Gly	Gln 415	Ala
:					103										120	
40	Gly	Val	Met	Gly	Phe	Pro	Gly	Pro	Lys	Gly	Ala	Ala	Gly	Glu	Pro	Gly
				420	,				425					430		
45	Lys	Ala	Gly	Glu	Arg	Gly	Val	Pro	Gly	Pro	Pro	Gly	Ala	Val	Gly	Pro
٠			435					440					445			
50	Ala	Gly 450	-	Asp	Gly	Glu	Ala 455	-	Ala	Gln	Gly	Pro		Gly	Pro	Ala

	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Glu	Gln	Gly	Pro	Ala	Gly	Ser	Pro	Gly
5	465					470					475					480
	Phe	Gln	Gly	Leu	Pro	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Glu	Ala	Gly	Lys
10					485					490					495	
		_	<b>-</b>				_				<b>~</b> 3	••-	<b>5</b>	<b>~</b> 3 -	_	
c.	Pro	Gly	Glu	500	GIÀ	Val	Pro	GIÀ	505	Leu	GIY	Ala	PIO	Gly 510	Pro	Ser
15				300					303							
	Gly	Ala	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Glu	Arg	Gly	Val	Gln	Gly
			515		•			520					525			
	Dro	Pro	Glv	Pro	Δla	Glv	Pro	Ara	Glv	Δla	Agn	G] v	Δla	Pro	Glv	Δan
	PIO	530	GLY	110	7124	O.J	535	*****	0.7		••	540			,	••••
25																
	Asp	Gly	Ala	Lys	Gly		Ala	Gly	Ala	Pro		Ala	Pro	Gly	Ser	
:	545					550					55 <b>5</b>					560
30	Glv	Ala	Pro	Gly	Leu	Gln	Gly	Met	Pro	Gly	Glu	Arg	Gly	Ala	Ala	Gly
	•			Ī	565		_			570					575	
ne																
35	Leu	Pro	Gly	Pro 580	Lys	Gly	Asp	Arg	Gly 585	Asp	Ala	Gly	Pro	Lys 590	Gly	Ala
				300					303					330		
40	Asp	Gly	Ser	Pro	Gly	Lys	Asp	Gly	Val	Arg	Gly	Leu	Thr	Gly	Pro	Ile
N.			595					600					605			
	Gly	Pro	Pro	Glv	Pro	בוג	Glv	Ala	Pro	G] v	Agn	Lva	Glv	Glu	Ser	Glv
45	GIY	610		GLY		ALU	615		110	O.J	p	620	U_j	-	-	07
•										( <b>-</b>						
50			Gly	Pro	Ala	_		Thr	Gly	Ala			Ala	Pro	Gly	
	625					630					635					640

	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Phe	Ala	Gly	Pro	Pro
5					645					650					655	
										•						
	Gly	Ala	Asp	Gly	Gln	Pro	Gly	Ala	Lys	Gly	Glu	Pro	Gly	Asp	Ala	Gly
40				660					665					670		
10																
	Ala	ГÀа	Gly	Asp	Ala	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro
			675					680					685			
15																
	Pro		Pro	Ile	Gly	Asn		Gly	Ala	Pro	Gly		Lys	Gly	Ala	Arg
		690					695					700				
20					_		_									
		Ser	Ala	Gly	Pro		Gly	Ala	Thr	Gly		Pro	Gly	Ala	Ala	Gly
	705					710					715					720
25	•	••- •	01	<b>D</b>	<b>5</b>	<b>a</b> 1	<b>-</b>	_					_	_		_
	Arg	vai	Gly	Pro		GIY	Pro	ser	GIĀ		ATA	GIÅ	Pro	Pro	_	Pro
<i>;</i>					725					730					735	
30	Dwo	Cly	Pro	בות	alv	Luc	C1	<b>~1</b>	~1··	T	<b>~1.</b> *	Dwa	N	<b>01</b>	<b>a</b> 1	m\
	PIO	Gry	PIO	740	GIY	пåз	GLU	GIŞ	745	гåя	GIĄ	PIO	ALG	750	GIU	THE
				710					743					750		
35	Glv	Pro	Ala	Glv	Ara	Pro	Glv	Glu	Val	Glv	Pro	Pro	Glv	Pro	Pro	Glv
	1		755	2	5		1	760		1			765			Cry
: 40	Pro	Ala	Gly	Glu	Lys	Gly	Ser	Pro	Gly	Ala	Asp	Gly	Pro	Ala	Gly	Ala
40		770					775					780			•	
	Pro	Gly	Thr	Pro	Gly	Pro	Gln	Gly	Ile	Ala	Gly	Gln	Arg	Gly	Val	Val
45	785					790					795					800
	Gly	Leu	Pro	Gly	Gln	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Leu	Pro	Gly
50					805					810					815	

	Pro	Ser	Gly	Glu	Pro	Gly	Lys	Gln	Gly	Pro	Ser	Gly	Ala	Ser	Gly	Glu
<b>.</b>				820					825					830		
5																
	Arg	Gly	Pro	Pro	Gly	Pro	Met	Gly	Pro	Pro	Gly	Leu	Ala	Gly	Pro	Pro
			835					840					845			
10																
	Gly	Glu	Ser	Gly	Arg	Glu	Gly	Ala	Pro	Gly	Ala	Glu	Gly	Ser	Pro	Gly
		850					855					860				
15																
	Arg	Asp	Gly	Ser	Pro	Gly	Ala	ГÀв	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro
	865					870					875					880
20																
	Ala	Gly	Pro	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Pro	Val
					885					890					895	
25																
	Gly	Pro	Ala	Gly	Lys	Ser	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro	Ala	Gly
				900					905					910		
30	Pro	Ala	Gly	Pro	Val	Gly	Pro	Ala	Gly	Ala	Arg	Gly	Pro	Ala	Gly	Pro
			915					920					925			
35	Gln	Gly	Pro	Arg	Gly	Asp	Lys	Gly	Glu	Thr	Gly	Glu	Gln	Gly	Asp	Arg
		930					935					940				
:																
40 <sup>-</sup>	Gly	Ile	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly		Gln	Gly	Pro	Pro	
	945					950					955					960
45	Pro	Pro	Gly	Ser		Gly	Glu	Gln	Gly		Ser	Gly	Ala	Ser		Pro
					965					970					975	
	,														_	
50	Ala	Gly	Pro	_	Gly	Pro	Pro	Gly		Ala	Gly	Ala	Pro	Gly	Lys	Asp
50				980					985					990		

	Gly	Leu	Asn	Gly	Leu	Pro	Gly	Pro	Ile	Gly	Pro	Pro	Gly	Pro	Arg	Gly
5			995					1000	0				100	5		
	Arg	Thr	Gly	Asp	Ala	Gly	Pro	Val	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro
10		1010	)				1019	5				1020	)			
	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Ser	Ala	Gly	Phe	Asp	Phe	Ser	Phe	Leu
15	1025	5				1030	)				103	5				1040
	Pro	Gln	Pro	Pro	Gln	Glu	Lys	Ala	His	qaA	Gly	Gly	Arg	Tyr	Tyr	Arg
20					1045	5				1050	)				105	5
20	212	Ara	Ser	Δla	T.011	Agn	Thr	) an	ጥረም	Cua	Dhe	Sar	Sar	The	G1	Tue
	ALU	~~3	<b>J</b> C2	1060		nop			106		2110	061	561	1070		пув
25																
	Asn	Cys	Cys		Arg	Gln	Leu	_		Asp	Phe	Arg	-	-	Leu	Gly
			1075	5				1080	)	•			1089	5		
30	Trp	Lys	Trp	Ile	His	Glu	Pro	Lys	Gly	Tyr	His	Ala	Asn	Phe	Cys	Leu
	_	1090					1099	-	_	-		1100			•	
35	Gly	Pro	Cys	Pro	Tyr	Ile	Trp	Ser	Leu	Asp	Thr	Gln	Tyr	Ser	Lys	
	1105	5				1110	)				1115	5				1120
40	Leu	Ala	Leu	Tyr	Asn	Gln	His	Asn	Pro	Glv	Ala	Ser	Ala	Ala	Pro	Cvs
				-	1125					1130					1139	-
45	Cys	Val	Pro	Gln	Ala	Leu	Glu	Pro	Leu	Pro	Ile	Val	Tyr	Tyr	Val	Gly
				1140	)				1145	5				1150	)	
	Δτα	Lva	Pro	ī.ve	Va 1	G1	G] ~	Lev	<b>c</b> e~	Aco	Ma+	TIA	r e v	A	Co-	C1
50	y	-70	1155		*41	GIU	<b>3111</b>	1160		Vali	MEL	116	1165	_	SEL	Cys

Lys	Сув	Sea
	1170	)

5

10

15

#### (2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3541 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

#### (ii) MOLECULE TYPE: cDNA

20

25

35

40

45

50

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGAAGGATT TCCATTTCCC AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT 60 TTCCGTGCCT GGCCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC ACCTGGTCCC CAAGGCTTCC AAGGTCCCCC TGGTGAGCCT GGCGAGCCTG GAGCTTCAGG TCCCATGGGT CCCCGAGGTC CCCCAGGTCC CCCTGGAAAG AATGGAGATG ATGGGGAAGC 240 TGGAAAACCT GGTCGTCCTG GTGAGCGTGG GCCTCCTGGG CCTCAGGGTG CTCGAGGATT 300 GCCCGGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTTCAGTG GTTTGGATGG 360 TGCCAAGGA GATGCTGGTC CTGCTGGTCC TAAGGGTGAG CCTGGCAGCC CTGGTGAAAA 420 TGGAGCTCCT GGTCAGATGG GCCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC 480 CCCTGGCCCT GCTGGTGCTC GTGGAAATGA TGGTGCTACT GGTGCTGCCG GGCCCCCTGG 540 TCCCACCGGC CCCGCTGGTC CTCCTGGCTT CCCTGGTGCT GTTGGTGCTA AGGGTGAAGC 600

660	AGCCTGGCCC	GTGCGTGGTG	TCCCCAGGGT	GCTCTGAAGG	GGGCCCCGAG	TGGTCCCCAA	
720	GACAGCCTGG	GGTGCTGATG	TGGAAACCCT	CTGGCCCTGC	GCTGGTGCTG	CCCTGGCCCT	
780	CTGGTGCCCG	CCTGGCTTCC	TGCTGGTGCT	CTCCTGGTAT	GCCAATGGTG	TGCTAAAGGT	
840	ACAGCGGTGA	CCCAAGGGTA	CCCTCCTGGT	GCCCCGGCGG	GGACCCCAGG	AGGCCCCTCT	
900	GCCCTGTTGG	GGAGAGCCTG	TGGTGCTAAG	AAGGAGACAC	CCTGGCAGCA	ACCTGGTGCT	
960	GAGGTGAACC	CGAGGAGCTC	GGAAGGAAAG	CTGCTGGAGA	CCCCCTGGCC	TGTTCAAGGA	
1020	GCCGTGGTTT	GGACCTGGTA	CGAGCGTGGT	GACCCCCTGG	GGCCTGCCCG	CGGACCCACT	
1080	GTTCTCCTGG	GGTGAACGTG	GGGTCCCGCT	CTGGTCCCAA	GATGGTGTTG	CCCTGGCGCA	
1140	CTGGTCTGCC	CCCGGTGAAG	AGCTGGTCGT	CTCCTGGTGA	CCCAAAGGAT	CCCCGCTGGC	
1200	AAACTGGCCC	CCTGATGGCA	CAGCCCTGGT	GAAGCCCTGG	GGTCTGACTG	TGGTGCCAAG	
1260	GTGCCCGTGG	GGCCCACCTG	CGGACCCCCA	ATGGTCGCCC	GCCGGTCAAG	CCCTGGTCCC	
1320	CCGGCAAGGC	GCTGGAGAGC	TAAAGGTGCT	TCCCTGGACC	GTGATGGGAT	TCAGGCTGGT	
1380	AAGATGGAGA	CCTGCTGGCA	CGCTGTCGGT	GACCCCCTGG	GGTGTTCCCG	TGGAGAGCGA	
1440	GTGAACAAGG	GGCGAGAGAG	TGGTCCCGCT	CTGGCCCTGC	CAGGGACCCC	GGCTGGAGCT	
1500	CAGGTGAAGC	GCTGGTCCTC	CCCTGGTCCT	TCCAGGGTCT	TCCCCCGGAT	CCCTGCTGGC	
1560	CCTCTGGAGC	GCCCTGGCC	AGACCTTGGC	GTGTTCCTGG	GGTGAACAGG	AGGCAAACCT	
1620	GTCCTGCTGG	GGTCCCCCTG	TGGTGTGCAA	CTGGCGAGCG	AGAGGTTTCC	AAGAGGCGAG	

ACCCCGAGGG	GCCAACGGTG	CTCCCGGCAA	CGATGGTGCT	AAGGGTGATG	CTGGTGCCCC	1680
TGGAGCTCCC	GGTAGCCAGG	GCGCCCTGG	CCTTCAGGĢA	ATGCCTGGTG	AACGTGGTGC	1740
AGCTGGTCTT	CCAGGGCCTA	AGGGTGACAG	AGGTGATGCT	GGTCCCAAAG	GTGCTGATGG	1800
CTCTCCTGGC	AAAGATGGCG	TCCGTGGTCT	GACCGGCCCC	ATTGGTCCTC	CTGGCCCTGC	1860
TGGTGCCCCT	GGTGACAAGG	GTGAAAGTGG	TCCCAGCGGC	CCTGCTGGTC	CCACTGGAGC	1920
TCGTGGTGCC	CCCGGAGACC	GTGGTGAGCC	TGGTCCCCCC	GGCCCTGCTG	GCTTTGCTGG	1980
CCCCCTGGT	GCTGACGGCC	AACCTGGTGC	TAAAGGCGAA	CCTGGTGATG	CTGGTGCCAA	2040
AGGCGATGCT	GGTCCCCCTG	GGCCTGCCGG	ACCCGCTGGA	CCCCTGGCC	CCATTGGTAA	2100
TGTTGGTGCT	CCTGGAGCCA	AAGGTGCTCG	CGGCAGCGCT	GGTCCCCCTG	GTGCTACTGG	2160
TTTCCCTGGT	GCTGCTGGCC	GAGTCGGTCC	TCCTGGCCCC	TCTGGAAATG	CTGGACCCC	2220
TGGCCCTCCT	GGTCCTGCTG	GCAAAGAAGG	CGGCAAAGGT	CCCCGTGGTG	AGACTGGCCC	2280
TGCTGGACGT	CCTGGTGAAG	TTGGTCCCCC	TGGTCCCCCT	GGCCCTGCTG	GCGAGAAAGG	2340
ATCCCCTGGT	GCTGATGGTC	CTGCTGGTGC	TCCTGGTACT	CCCGGGCCTC	AAGGTATTGC	2400
TGGACAGCGT	GGTGTGGTCG	GCCTGCCTGG	TCAGAGAGGA	GAGAGAGGCT	TCCCTGGTCT	2460
TCCTGGCCCC	TCTGGTGAAC	CTGGCAAACA	AGGTCCCTCT	GGAGCAAGTG	GTGAACGTGG	2520
TCCCCCCGGT	CCCATGGGCC	CCCCTGGATT	GGCTGGACCC	CCTGGTGAAT	CTGGACGTGA	2580
GGGGGCTCCT	GCTGCCGAAG	GTTCCCCTGG	ACGAGACGGT	TCTCCTGGCG	CCAAGGGTGA	2640

. 20

i 

50-

CCGTGGTGAG	ACCGGCCCCG	CTGGACCCCC	TGGTGCTCCT	GGTGCTCCTG	GTGCCCCTGG	2700
CCCCGTTGGC	CCTGCTGGCA	AGAGTGGTGA	TCGTGGTGAG	ACTGGTCCTG	CTGGTCCCGC	2760
CGGTCCCGTC	GGCCCCGCTG	GCGCCCGTGG	CCCCGCCGGA	CCCCAAGGCC	CCCGTGGTGA	2820
CAAGGGTGAG	ACAGGCGAAC	AGGGCGACAG	AGGCATAAAG	GGTCACCGTG	GCTTCTCTGG	2880
CCTCCAGGGT	CCCCTGGCC	CTCCTGGCTC	TCCTGGTGAA	CAAGGTCCCT	CTGGAGCCTC	2940
TGGTCCTGCT	GGTCCCCGAG	GTCCCCCTGG	CTCTGCTGGT	GCTCCTGGCA	AAGATGGACT	3000
CAACGGTCTC	CCTGGCCCCA	TTGGGCCCCC	TGGTCCTCGC	GGTCGCACTG	GTGATGCTGG	3060
TCCTGTTGGT	ccccccccc	CTCCTGGACC	TCCTGGTCCC	CCTGGTCCTC	CCAGCGCTGG	3120
TTTCGACTTC	AGCTTCCTCC	CCCAGCCACC	TCAAGAGAAG	GCTCACGATG	GTGGCCGCTA	3180
CTACCGGGCT	AGATCTGCCC	TGGACACCAA	CTATTGCTTC	AGCTCCACGG	AGAAGAACTG	3240
CTGCGTGCGG	CAGCTGTACA	TTGACTTCCG	CAAGGACCTC	GGCTGGAAGT	GGATCCACGA	3300
GCCCAAGGGC	TACCATGCCA	ACTTCTGCCT	CGGGCCCTGC	CCCTACATTT	GGAGCCTGGA	3360
CACGCAGTAC	AGCAAGGTCC	TGGCCCTGTA	CAACCAGCAT	AACCCGGGCG	CCTCGGCGGC	3420
GCCGTGCTGC	GTGCCGCAGG	CGCTGGAGCC	GCTGCCCATC	GTGTACTACG	TGGGCCGCAA	3480
GCCCAAGGTG	GAGCAGCTGT	CCAACATGAT	CGTGCGCTCC	TGCAAGTGCA	GCTGATCTAG	3540
Α .						3541

:	(2)	INFOR	MATI	ON F	OR S	EQ 1	טא ט.	):10:									
5		(;)	S FOT	IENCE	з сна	ריז מים	ERTS	የተተርያ	٠.								
		(1)	_							la.	•						
4.					igth : _				acic	15							
10			(B)	TYI	PE: a	ımınc	acı	ıa									
			(C)	STF	LANDE	EDNES	S: £	ingl	.e								
			(D)	TOI	OLOG	Υ: υ	ınkno	wn									
S.																	
15		(ii)	MOLE	CUL	TYP	PE: p	epti	ide									
		(xi)	SEQU	JENCI	E DES	CRI	PTION	1: SI	EQ II	NO:	10:						
20																	
•		Gln	Leu	Ser	Tyr	Gly	Tyr	Asp	Glu	Lys	Ser	Thr	Gly	Gly	Ile	Ser	Val
		1				5					10					15	
•																	
25		Pro	Gly	Pro	Met	Gly	Pro	Ser	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Pro	Pro
					20					25					30		
:																	
30		Glv	Ala	Pro	Gly	Pro	Gln	Gly	Phe	Gln	Gly	Pro	Pro	Gly	Glu	Pro	Gly
		•		35	_			_	40		_			45			-
:-		<b>61.</b> 1	Dro	alv	פות	Cor	G1 v	Dro	Mat	Gly	Pro	λνα	Glv	Pro	Pro	Gly	Pro
35		GIU		GIY	MIG	Ser	GLY		MEC	GIY	FIU	AL 9		710		GLY	110
			50					55					60				
											_	_			_		
		Pro	Gly	Lys	Asn	Gly	Asp	Asp	Gly	Glu	Ala		Lys	Pro	Gly	Arg	
40		65					70					75					80
		Gly	Glu	Arg	Gly	Pro	Pro	Gly	Pro	Gln	Gly	Ala	Arg	Gly	Leu	Pro	Gly
45						85					90					95	
		Thr	Ala	Gly	Leu	Pro	Gly	Met	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly	Leu
					100		_		-	105		,			110		
50																	

١.

ÿ.,

	Asp	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Glu	Pro
5			115					120					125			
	Gly	Ser	Pro	Gly	Glu	Asn	Gly	Ala	Pro	Gly	Gln	Met	Gly	Pro	Arg	Gly
10		130					135					140				
m.	Leu	Pro	Gly	Glu	Arg	Gly	Arg	Pro	Gly	Ala	Pro	Gly	Pro	Ala	Gly	Ala
15	145					150					155					160
	Arg	Gly	Asn	Asp		Ala	Thr	Gly	Ala		Gly	Pro	Pro	Gly		Thr
					165					170					175	
20										_						
•	Gly	Pro	Ala		Pro	Pro	Gly	Phe		Gly	Ala	Val	Gly		Lys	Gly
				180					185					190		
25			<b>43</b>		<b>~</b> 1	<b>01</b>	<b>5</b>		<b>a1</b>	<b>a</b>	<b>a</b> 1	<b>a</b> 1	D	<b>71</b> -	al	17- 3
	Glu	Ala		Pro	Gin	GIY	Pro		GIĀ	ser	GIU	GIY	205	GIN	GIÀ	vai
			195					200					203			
30	•	Gly	<b>~</b> 1	Dwa	<u>ما</u>	D=0	D=0	<i>α</i> 1	Dro	ח ה	Gly.	λla	al a	GIV.	Pro	בות
	Arg	_	GIU	PIO	GIY	PIO	215	GIŞ	PIO	мта	GIY	220	AIG	GIY	PLO	AIG
		210					213								•	
35	Glv	) an	Pro	Glv	ั้มโล	Asn	G] v	Gln	Pro	Glv	Ala	Lvs	Glv	Ala	Asn	Gly
	225					230		<b></b>		1	235		2			240
•	227															
40	Ala	Pro	Gly	Ile	Ala	Gly	Ala	Pro	Gly	Phe	Pro	Gly	Ala	Arg	Gly	Pro
			Ī		245				_	250					255	
45	Ser	Gly	Pro	Gln	Gly	Pro	Gly	Gly	Pro	Pro	Gly	Pro	Lys	Gly	Asn	Ser
				260					265					270		
50	Gly	Glu	Pro	Gly	Ala	Pro	Gly	Ser	Lys	Gly	Asp	Thr	Gly	Ala	Lys	Gly
50			275					280					285			
							•									

<i>)</i> .	Glu	Pro	Gly	Pro	Val	Gly	Val	Gln	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Glu
5		290					295					300				
۸.		Gly	Lys	Arg	Gly		Arg	Gly	Glu	Pro		Pro	Thr	Gly	Leu	
10	305					310					315					320
•	Gly	Pro	Pro	Gly	Glu 325	Arg	Gly	Gly	Pro	Gly 330	Ser	Arg	Gly	Phe	Pro	Gly
15																
	Ala	Asp	Gly	Val 340	Ala	Gly	Pro	Lys	Gly 345	Pro	Ala	Gly	Glu	Arg 350	Gly	Ser
20																
=	Pro	Gly	Pro 355	Ala	Gly	Pro	Lys	Gly 360	Ser	Pro	Gly	Glu	<b>Ala</b> 365	Gly	Arg	Pro
25			- "		_	_										
es.	GIÀ	370	Ala	GIY	Leu	Pro	375 Gly	Ala	Lys	Gly	Leu	380	Gly	Ser	Pro	Gly
30	Ser	Pro	Glv	Pro	Agn	Glv	Tava	Thr	Glv	Pro	Pro	Glv	Dro	פות	Gly	<b>~1</b> ~
	385		<b>-</b>		p	390	-,0		,		395	O.J		ALG	GLY	400
35	Asp	Gly	Arg	Pro		Pro	Pro	Gly	Pro	Pro	Gly	Ala	Arg	Gly	Gln	Ala
					405					410					415	
40	Gly	Val	Met		Phe	Pro	Gly	Pro	Lys	Gly	Ala	Ala	Gly	Glu	Pro	Gly
				420					425					430		
45	Lys	Ala		Glu	Arg	Gly	Val		Gly	Pro	Pro	Gly		Val	Gly	Pro
			435					440					445			
50	Ala	Gly 450	Lys	Asp	Gly	Glu		Gly	Ala	Gln	Gly		Pro	Gly	Pro	Ala
4		*30					455					460				

	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Glu	Gln	Gly	Pro	Ala	Gly	Ser	Pro	Gly
5	465					470				•	475					480
s .	Phe	Gln	Gly	Leu	Pro 485	Gly	Pro	Ala	Gly	Pro 490	Pro	Gly	Glu	Ala	Gly 495	Lys
.10																
í	Pro	Gly	Glu	Gln 500	Gly	Val	Pro	Gly	Asp 505	Leu	Gly	Ala	Pro	Gly 510	Pro	Ser
15																
	Gly	Ala	Arg 515	Gly	Glu	Arg	Gly	Phe 520	Pro	Gly	Glu	Arg	Gly 525	Val	Gln	Gly
20				_			_				_			_		_
	Pro	9ro 530	Gly	Pro	Ala	GIÀ	Pro 535	Arg	GIÀ	Ala	Asn	G19 540	Ala	Pro	Gly	Asn
25	Asp	Gly	Ala	Lys	Gly	_	Ala	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ser	
	545					550					555					560
30	Gly	Ala	Pro	Gly	Leu 565	Gln	Gly	Met	Pro	Gly 570	Glu	Arg	Gly	Ala	Ala 575	Gly
٠.					303					3,0			ı		3.3	
<b>35</b> '	Leu	Pro	Gly	Pro 580	Lys	Gly	Asp	Arg	Gly 585	Asp	Ala	Gly	Pro	Lys 590	Gly	Ala
40· ·	Asp	Gly	Ser 595	Pro	Gly	Lys	Asp	Gly 600	Val	Arg	Gly	Leu	Thr 605	Gly	Pro	Ile
45	Gly	Pro 610		Gly	Pro	Ala	Gly 615		Pro	Gly	Ąsp	Lys 620		Glu	Ser	Gly
50	Pro 625		Gly	Pro	Ala	Gly 630		Thr	Gly	Ala	Arg		Ala	Pro	Gly	Asp 640

	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Phe	Ala	Gly	Pro	Pro
5					645					650					655	
•	Gly	Ala	Asp	Gly	Gln	Pro	Gly	Ala	Lys	Gly	Glu	Pro	Gly	Asp	Ala	Gly
				660					665					670		
10																
	Ala	Lys	_	Asp	Ala	Gly	Pro		Gly	Pro	Ala	Gly		Ala	Gly	Pro
			675					680	٠				685			
15	Pro	Glv	Pro	Tle	Gly	Δsn	va1	Glv	Δla	Pro	Glv	Ala	Lvs	Glv	Ala	Ara
•	PIO	690			UL,		695	O.J			o.,	700		O.J	7124	y
20																
20	Gly	Ser	Ala	Gly	Pro	Pro	Gly	Ala	Thr	Gly	Phe	Pro	Gly	Ala	Ala	Gly
	705					710					715					720
25												=				
•	Arg	Val	Gly	Pro	Pro	Gly	Pro	Ser	Gly		Ala	Gly	Pro	Pro		Pro
					725					730					735	
30	Pro	Gly	Pro	Ala	Gly	Lys	Glu	Gly	Gly	Lys	Gly	Pro	Arg	Gly	Glu	Thr
				740					745					750		
35	Gly	Pro	Ala	Gly	Arg	Pro	Gly	Glu	Val	Gly	Pro	Pro		Pro	Pro	Gly
•			755					760					765			
	Pro	Δla	Glv	Glu	Lys	Glv	Ser	Pro	Glv	Ala	asp	Glv	Pro	Ala	Glv	Ala
40	110	770	<b>U</b> -1	-	-,-		775		1			780			2	
·-	Pro	Gly	Thr	Pro	Gly	Pro	Gln	Gly	Ile	Ala	Gly	Gln	Arg	Gly	Va1	Val
<b>45</b>	785					790				·	795					800
ŕ		_	_	<b>~</b> 1	<b></b> 1		<b>6</b> 3	<b>a</b> 3	•	<b>~</b> 3	nt-	<b>n</b>	01	•	Door	<b>03</b>
50	Gly	Leu	Pro	GΙΆ	Gln 805	Arg	GIY	GIU	arg	810		Pro	GIĄ	Leu	Pro 815	GΤÅ
					503					010						

	Pro	Ser	Gly	Glu	Pro	Gly	Lys	Gln	Gly	Pro	Ser	Gly	Ala	Ser	Gly	Glu
.; <b>5</b>				820					825					830		
٠	_				<b>-1</b>	<b>D</b>	14-4	<b>~</b> 1	D		<b>a</b> 1	<b>T</b>	21.	<b>61</b>	D	<b>.</b>
	Arg	GIÀ	835	Pro	GIĄ	PTO	met	61y 840	PIO	PIO	GIÀ	Leu	845	GIY	PTO	Pro
.10			633					040		•			013			
	Gly	Glu	Ser	Gly	Arg	Glu	дÌу	Ala	Pro	Gly	Ala	Glu	Gly	Ser	Pro	Gly
.15	_	850					855					860				
•	Arg	Asp	Gly	Ser	Pro	Gly	Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro
20	865					870					875					880
										_			_			
	Ala	Gly	Pro	Pro	_	Ala	Pro	Gly	Ala		GTA	Ala	Pro	GΙΆ		Val
25					885					890					895	
	Glv	Pro	Ala	Gly	Lys	Ser	Gly	Asp	Arq	Gly	Glu	Thr	Gly	Pro	Ala	Gly
	•			900			-	_	905					910		_
30																
	Pro	Ala	Gly	Pro	Val	Gly	Pro	Ala	Gly	Ala	Arg	Gly	Pro	Ala	Gly	Pro
			915					920					925			
35															_	
	Gln	_	Pro	Arg	Gly	Asp		Gly	Glu	Thr	Gly		Gln	Gly	qaA	Arg
		930					935					940				
40	Glv	Ile	Lvs	Glv	His	Arg	Glv	Phe	Ser	Gly	Leu	Gln	Gly	Pro	Pro	Gly
	945		•	•		950	•			•	955		-			960
45	Pro	Pro	Gly	Ser	Pro	Gly	Glu	Gln	Gly	Pro	Ser	Gly	Ala	Ser	Gly	Pro
٠,					965					970					975	
															_	
50	Ala	Gly	Pro		Gly	Pro	Pro	Gly		Ala	Gly	Ala	Pro		Lys	Asp
				980					985					990		

	Gly	Leu	Asn	Gly	Leu	Pro	Gly	Pro	Ile	Gly	Pro	Pro	Gly	Pro	Arg	Gly
5			995					1000	)				1005	•		
:	Arg		_	Asp	Ala	Gly	Pro		Gly	Pro	Pro	Gly		Pro	Gly	Pro
10		1010	,				101:	,				102				
<b>:</b>	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Ser	Ala	Gly	Phe	Asp	Phe	Ser	Phe	Leu
15	1025	5				1030	)				1035	5				1040
	Pro	Gln	Pro	Pro			Lys	Ala	His	Asp		Gly	Arg	Tyr		
20					104	5				1050	J				105	•
	Ala	Arg	Ser	Asp	Glu	Ala	Ser	Gly	Ile	Gly	Pro	Glu	Val	Pro	Asp	Asp
				106	0				106	5				1070	)	
25																
•	Arg	Asp	Phe	Glu	Pro	Ser	Leu	Gly	Pro	Val	Сув	Pro	Phe	Arg	Cys	Gln
ž.			107	5				108	0				108	5		
30																
	Сув	His	Leu	Arg	Val	Val	Gln	Сув	Ser	Asp	Leu	Gly	Leu	Asp	Lys	Val
o-		109	0				109	5				110	0			
35	Pro	Lys	Asp	Leu	Pro	Pro	Asp	Thr	Thr	Leu	Leu	Asp	Leu	Gln	Asn	Asn
71	110	5				111	0				111	5				1120
40	Lys	Ile	Thr	Glu	Ile	Lys	Asp	Gly	Asp	Phe	Lys	Asn	Leu	Lys	Asn	Leu
					112	5				113	0				113	5
45	His	Ala	Leu	Ile	Leu	Val	Asn	Asn	Lys	Ile	Ser	Lys	Val	Ser	Pro	Gly
				114	0				114	5				115	0	
50	Ala	Phe	Thr	Pro	Leu	Val	Lys	Leu	Glu	Arg	Leu	Tyr	Leu	Ser	Lys	Asn
			115	5				116	0				116	5		
1																

	Gln	Leu	ГÀв	Glu	Leu	Pro	Glu	Lys	Met	Pro	Lys	Thr	Leu	Gln	Glu	Leu
5		1170	)				1175	,				1180	)			
	Arg	Ala	His	Glu	Asn	Glu	Ile	Thr	Lys	Val	Arg	Lys	Val	Thr	Phe	Asn
10	1185	5				1190	)				1195	5				1200
	Gly	Leu	Asn	Gln	Met	Ile	Val	Ile	Glu	Leu	Gly	Thr	Asn	Pro	Leu	Lys
15					1205	5				1210	)				1215	5
	ser	Ser	Gly	Ile	Glu	Asn	Gly	Ala	Phe	Gln	Gly	Met	Lys	Lys	Leu	Ser
20				1220	)				1225	5				1230	)	
20																
	Tyr	Ile	Arg	Ile	Ala	Asp	Thr	Asn	Ile	Thr	Ser	Ile	Pro	Gln	Gly	Leu
			1235	5 .				1240	)				1245	5		
25																
	Pro		Ser	Leu	Thr	Glu			Leu	Asp	Gly			Ile	Ser	Arg
		1250	0				125	5				1260	)			
30		_			_		_	~3	-	•		¥		<b>7</b>	•	<b>~</b> 3
,		_	Ala	Ala	ser			GIY	Leu	Asn	127		Ala	гÀв	Leu	Gly 1280
	126	•				127	U				127	•				1260
35	T.411	ger.	Phe	λen	Ser	Tle	Ser	Ala	Val	Asp	Asn	Glv	Ser	Leu	Ala	Asn
	Deu	362	****		128		-			129		1			129	
40 .	Thr	Pro	His	Leu	Arg	Glu	Leu	His	Leu	Asp	Asn	Asn	Lys	Leu	Thr	Arg
				130	0				130	5				131	0	
45	Val	Pro	Gly	Gly	Leu	Ala	Glu	His	Lys	Tyr	Ile	Gln	Val	Val	Tyr	Leu
			131	5				132	0				132	5		
50	His	Asn	Asn	Asn	Ile	Ser	Val	Val	Gly	Ser	Ser	Asp	Phe	Сув	Pro	Pro
		133	0				133	5				134	0			

4.	Gly His Asn	Thr Lys Lys	Ala Ser Tyr S	er Gly Val Ser	Leu Phe Ser
5	1345	1350	)	1355	1360
	Asn Pro Val		Glu Ile Gln F	ro Ser Thr Phe	
10		1365	· 1	.370	1375
ťe	Tyr Val Arg		Gln Leu Gly A	an Tyr Lys	
15		1380	1385		
•	(2) INFORMATION	FOR SEC ID NO	)·11·		•
	(2) Intoldation				
20	(i) SEQUENC	E CHARACTERIS	STICS:		
•	(A) LE	NGTH: 1107 an	mino acids		
	(B) TY	PE: amino aci	id		
25	(C) ST	RANDEDNESS: S	single	,	
	(D) TO	POLOGY: unkno	own		
30	(ii) MOLECUI	E TYPE: pept:	ide		
	(wi) CEOUTENO	'E DECCETOTA	N: SEQ ID NO:	11.	
	(XI) SEQUENC	.E DESCRIPTION	N: SEQ ID NO.		
35	Gln Leu Sei	Tyr Gly Tyr	Asp Glu Lys S	Ser Thr Gly Gly	Ile Ser Val
	1	5		10	15
٠.					
40	Pro Gly Pro	Met Gly Pro	Ser Gly Pro	Arg Gly Leu Pro	Gly Pro Pro
	•	20	25		30
•					
45		Gly Pro Gln		Gly Pro Pro Gly	Glu Pro Gly
	35		40	45	
	Glu Pro Glu	/ Ala Ser Glv	Pro Met Glv	Pro Arg Gly Pro	Pro Glv Pro
50	50		55	60	,

÷ "	Pro	Gly	Lys	Asn	Gly	Asp	Asp	Gly	Glu	Ala	Gly	Lys	Pro	Gly	Arg	Pro
5	65					70					75					80
5	Gly	Glu	Arg	Gly	Pro	Pro	Gly	Pro	Gln	Gly	Ala	Arg	Gly	Leu	Pro	Gly
10					85					90					95	
	Thr	Ala	Gly	Leu	Pro	Gly	Met	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly	Leu
15				100			<b>.</b>		105					110		
	Asp	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Glu	Pro
20			115					120					125			
	<b>61</b>		Dwa	<b>~1</b>	<b>~1</b>	<b>3</b> ~ =	~1··	330	Dwa	~1·•	<b>~</b> 1~	Mot	<i>~</i> 1	Dwa	7	<b>61</b>
	GIY	130	PIO	GIY	Giu	ASII	135	AIG	PIO	GIA	GIII	140	GIY	PLO	AIG	Gly
25		130					133							1		
	Leu	Pro	Gly	Glu	Arg	Gly	Arg	Pro	Gly	Ala	Pro	Gly	Pro	Ala	Gly	Ala
ž.	145					150					155					160
30																
	Arg	Gly	Asn	Asp	Gly	Ala	Thr	Gly	Ala	Ala	Gly	Pro	Pro	Gly	Pro	Thr
6 ·					165					170					175	
35		_		<b>~</b> 3			-1		_	<b>~1</b>		··· •	<b>a</b> 1			
	GIÀ	Pro	Ala	180	Pro	Pro	GIA	Phe	Pro 185	GIÀ	Ala	vaı	GTÅ	190	Lys	Gly
				180					103					130		
40	Glu	Ala	Gly	Pro	Gln	Gly	Pro	Arg	Gly	Ser	Glu	Gly	Pro	Gln	Gly	Val
			195			_		200	_				205		_	
45	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Ala	Gly	Pro	Ala
		210					215					220	٠			
				_									_	_		
50		Asn	Pro	Gly	Ala		Gly	Gln	Pro	Gly		Lys	Gly	Ala	Asn	Gly
	225					230					235					240

	Ala	Pro	Gly	Ile	Ala	Gly	Ala	Pro	Gly	Phe	Pro	Gly	Ala	Arg	Gly	Pro
5					245					250					255	
	Cor	Glv	Pro	Gln	Gly	Pro	Glv	g] v	Pro	Pro	Glv	Pro	Lvs	Glv	Asn	Ser
	361	O.J		260	02,		,	<b>02</b> 7	265		1		-,-	270		
10																
	Gly	Glu	Pro	Gly	Ala	Pro	Gly	Ser	Lys	Gly	Aap	Thr	Gly	Ala	Lys	Gly
			275					280					285			
15																
	Glu	Pro	Gly	Pro	Val	Gly		Gln	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Glu
		290					295					300				
20				•	<b>a</b> 3	•••	•	<b>41</b>	<b>a1</b>	<b>D</b>	<b>a</b> 1	D-10	س <i>ا</i>	<b>61</b>	<b>T</b>	D
		GIÀ	гув	Arg	Gly		Arg	GIA	GIU	PIO		Pro	Int	GIÀ	ren	320
	305					310					315					320
25	Glv	Pro	Pro	Glv	Glu	Arg	Glv	Glv	Pro	Glv	Ser	Arq	Glv	Phe	Pro	Glv
	<b>0</b> -7			,	325			2		330			•		335	4
30	Ala	Asp	Gly	Val	Ala	Gly	Pro	Lys	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Ser
				340					345					350		
•																
35	Pro	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Ser	Pro	Gly	Glu		Gly	Arg	Pro
			355					360					365			
,		<b>~3</b>	•••	<b>61</b>	•	<b>5</b>	<b>~</b> 1	•••	<b>T</b>	<b>a</b> 1	<b>7</b>	mh	<b>~1</b>	Ca	Dw.s	<b>01</b>
<b>40</b> <sup>1</sup>	GTÅ		Ala	GIY	Leu	PIO		Ala	гув	GIŞ	Leu	380	GIY	ser	PIO	GIY
		370					375					300				
	Ser	Pro	Gly	Pro	Asp	Gly	Lys	Thr	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Gln
45	385		-		_	390	-				395					400
	Asp	Gly	Arg	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Ala	Arg	Gly	Gln	Ala
50					405					410					415	

	Gly	Val	Met	Gly	Phe	Pro	Gly	Pro	Lys	Gly	Ala	Ala	Gly	Glu	Pro	Gly
5				420					425					430		
	Lys	Ala	Gly	Glu	Arg	Gly	Val	Pro	Gly	Pro	Pro	Gly	Ala	Val	Gly	Pro
10			435					440					445			
•	Ala	Gly	Lys	Asp	Gly	Glu	Ala	Gly	Ala	Gln	Gly	Pro	Pro	Gly	Pro	Ala
15		450					455					460				
	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Glu	Gln	Gly	Pro	Ala	Gly	Ser	Pro	Gly
20	465		•			470					475					480
	Phe	Gln	Gly	Leu	Pro	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Glu	Ala	Gly	Lys
25					485					490					495	
	Pro	Gly	Glu	Gln	Gly	Val	Pro	Gly	Asp	Leu	Gly	Ala	Pro	Gly	Pro	Ser
				500					505					510		
30	Gly	Ala	Ārg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Glu	Arg	Gly	Val	Gln	Gly
			515					520					525			
35	Pro	Pro	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Ala	Asn	Gly	Ala	Pro	Gly	Asn
		530					535					540				
÷ 40	•	<b>63</b>		Lys	<b>~1</b>	7.50	21-	<b>63</b>	<b>31</b> -	Dro	G1v	77.0	Pro	alv	Sar	al n
	545	-	Ald	nys	GIĀ	550	Ala	GIÅ	ALA	PLO	555	ALG	PLO	GIY	Ser	560
45	GJY	Ala	Pro	Gly	Leu 565	Gln	Gly	Met	Pro	Gly 570	Glu	Arg	Gly	Ala	Ala 575	Gly
50	Leu	Pro	Gly	Pro	Lys	Gly	Asp	Arg	Gly	Asp	Ala	Gly	Pro	Lys	Gly	Ala
				580					585					590		

_	Asp	Gly	Ser	Pro	Gly	Lys	Asp	Gly	Val	Arg	Gly	Leu	Thr	Gly	Pro	Ile
5			595					600					605			
	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Pro	Gly	Asp	Lys	Gly	Glu	Ser	Gly
10		610					615					620				
		Ser	Gly	Pro	Ala		Pro	Thr	Gly	Ala		Gly	Ala	Pro	Gly	Asp
15	625					630					635					640
	A Y C	Glv	Glu	Pro	G) v	Pro	Pro	Glv	Pro	בות	Gly	Dhe	פומ	Gly	Dwa	Dwa
	Æg	ULY	<b>01</b> 4		645		110	GLY	110	650	Gry	FILE	ΛIα	Gry	655	PIO
20					013					030					655	
	Gly	Ala	Asp	Gly	Gln	Pro	Gly	Ala	Lys	Gly	Glu	Pro	Gly	qaA	Ala	Gly
				660					665					670		
25											.,					
	Ala	Lys	Gly	Asp	Ala	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro
,			675					680					685			
30																
	Pro	Gly	Pro	Ile	Gly	Asn	Val	Gly	Ala	Pro	Gly	Ala	Lys	Gly	Ala	Arg
•		690					695					700				
35 <sup>†</sup>	_						_	_								
		Ser	Ala	Gly	Pro		Gly	Ala	Thr	Gly		Pro	Gly	Ala	Ala	_
:-	705					710					715					720
40	Ara	Val	G] v	Pro	Pro	Glv	Pro	Ser.	Glv	) en	בומ	alv.	Bro	Pro	C1	Dwa
•	g	• • • • • • • • • • • • • • • • • • • •	O.J		725	023	110	Der	GLY	730	ALG	GIY	PIO	PIO	735	PIO
										,,,,					,,,,	
45	Pro	Gly	Pro	Ala	Gly	Lys	Glu	Gly	Gly	Lys	Gly	Pro	Arg	Gly	Glu	Thr
				740				_	745	_	_			750		
50	Gly	Pro	Ala	Gly	Arg	Pro	Gly	Glu	Val	Gly	Pro	Pro	Gly	Pro	Pro	Gly
•			755					760					765			
<b>.</b>																

	Pro	Ala	Gly	Glu	Lys	Gly	Ser	Pro	Gly	Ala	qeA	Gly	Pro	Ala	Gly	Ala
5		770					775					780				
٠,																
•	Pro	Gly	Thr	Pro	Gly	Pro	Gln	Gly	Ile	Ala	Gly	Gln	Arg	Gly	Val	Val
10	785					790					795					800
	Glv	Leu	Pro	Glv	Gln	Arq	Glv	Glu	Arq	Gly	Phe	Pro	Gly	Leu	Pro	Glv
	027				805	- 3				810			•		815	3
.15					•••					•						
	D	C	G1	G1.1	Dro	C1.	Tva	Cln.	GT v	D×0	Sar	al v	פומ	Ca~	Gly	C1
	Pro	Ser	GIY		PLO	GIY	пåа	GIII	-	PLO	261	GIŞ	ATG		GIŞ	GIU
20				820					825					830		
					_							_				
	Arg	Gly		Pro	Gly	Pro	Met	Gly	Pro	Pro	Gly	Leu		GIA	Pro	Pro
			835					840					845			
25																
	Gly	Glu	Ser	Gly	Arg	Glu	Gly	Ala	Pro	Gly	Ala	Glu	Gly	Ser	Pro	Gly
		850					855					860				
30																
	Arg	Asp	Gly	Ser	Pro	Gly	Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro
•	865					870					875					880
95																
35	Ala	Gly	Pro	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Pro	Val
\$-					885					890					895	
40 <sup>-</sup>	Glv	Pro	Ala	Glv	Lvs	Ser	Glv	Asp	Arg	Glv	Glu	Thr	Glv	Pro	Ala	Glv
	027			900	-2-		2		905					910		2
				,,,,					,							
45	D~~	<b>11</b> -	رد ا در.	Dro	V=1	Gl v	Dra	- ומ	Glav	פומ	Arc	Glv	Dro	αſα	Gly	Dro
	PIO	ALG	_		vai	GIY	PIO		GIY	MIG	мy	GLY			GLY	210
			915					920					925			
			_	_		_	_					<b></b> -			_	
50	Gln	-		Arg	Gly	Asp	-	_	Glu	Thr	Gly		Gln	Gly	Asp	Arg
		930					935					940				

55

.:

	Gly	Ile	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly	Leu	Gln	Gly	Pro	Pro	Gly
5	945					950					955					960
			•													
	Pro	Pro	Gly	Ser	Pro	Gly	Glu	Gln	Gly		Ser	Gly	Ala	Ser		Pro
10					965					970					975	
										_			_			
	Ala	Gly	Pro		Gly	Pro	Pro	Gly		Ala	Gly	Ala	Pro		Lys	Asp
15				980					985					990		
		_		<b>~</b> 3	•		<b>a</b> 1	<b>-</b>	-1.	g1	D	7	<b>61</b>	D	•	<b>01</b>
	Gly	Leu		GIÀ	Leu	Pro	GTÀ			GIĀ	PIO	PIO			Arg	Gly
20			995					1000	)				100	•		
	•	mb	a1	3 an	77-	<b>~1</b>	Dwo	v-1	<b>61</b> 14	Dro	Dro	Glv	Pro	Pro	Gly	Pro
	Arg	1010	-	wah	Ala	GIY	101		GLY	FIO	PIO	102		710	GLY	FIO
25		101(	,				101.	,				102	•			
	Pro	Glv	Pro	Pro	Glv	Pro	Pro	Ser	Ala	Gly	Phe	Asp	Phe	Ser	Phe	Leu
_	1025	_			•	103				•	103					1040
30																
	Pro	Gln	Pro	Pro	Gln	Glu	Lys	Ala	His	Asp	Gly	Gly	Arg	Tyr	Tyr	Arg
					104	5				105	0				105	5
35																
	Ala	Arg	Ser	Pro	Lys	Asp	Leu	Pro	Pro	Asp	Thr	Thr	Leu	Leu	Asp	Leu
				106	0				106	5				107	0	
40																
••	Gln	Asn	Asn	Lys	Ile	Thr	Glu	Ile	Lys	Asp	Gly	Asp	Phe	Lys	Asn	Leu
			107	5				108	0				108	5		
45 <sup>.</sup>	Lys	Asn	Leu	His	Ala	Leu	Ile	Leu	Val	Asn	Asn	Lys	Ile	Ser	Lys	Val
		109	0			•	109	5				110	0			
50	Ser	Pro	Gly	•												
	110	5														
1																

55·

#### (2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4167 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

25

30

35

40

45

50

5

10

15

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC TGGCCCCATG 60 GGTCCCTCTG GTCCTCGTGG TCTCCCTGGC CCCCTGGTG CACCTGGTCC CCAAGGCTTC 120 CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT 180 CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT 240 GGTGAGCGTG GGCCTCCTGC GCCTCAGGGT GCTCGAGGAT TGCCCGGAAC AGCTGGCCTC 300 CCTGGAATGA AGGGACACAG AGGTTTCAGT GGTTTGGATG GTGCCAAGGG AGATGCTGGT 360 CCTGCTGGTC CTAAGGGTGA GCCTGGCAGC CCTGGTGAAA ATGGAGCTCC TGGTCAGATG 420 GGCCCCGTG GCCTGCCTGG TGAGAGAGGT CGCCCTGGAG CCCCTGGCCC TGCTGGTGCT 480 CGTGGAAATG ATGGTGCTAC TGGTGCTGCC GGGCCCCCTG GTCCCACCGG CCCCGCTGGT 540 CCTCCTGGCT TCCCTGGTGC TGTTGGTGCT AAGGGTGAAG CTGGTCCCCA AGGGCCCCGA 600

GGCTCTGAAG	GTCCCCAGGG	TGTGCGTGGT	GAGCCTGGCC	CCCCTGGCCC	TGCTGGTGCT	660
GCTGGCCCTG	CTGGAAACCC	TGGTGCTGAT	GGACAGCCȚG	GTGCTAAAGG	TGCCAATGGT	720
GCTCCTGGTA	TTGCTGGTGC	TCCTGGCTTC	CCTGGTGCCC	GAGGCCCCTC	TGGACCCCAG	780
GGCCCCGGCG	GCCCTCCTGG	TCCCAAGGGT	AACAGCGGTG	AACCTGGTGC	TCCTGGCAGC	840
AAAGGAGACA	CTGGTGCTAA	GGGAGAGCCT	GGCCCTGTTG	GTGTTCAAGG	ACCCCCTGGC	900
CCTGCTGGAG	AGCAAGGAAA	GCGAGGAGCT	CGAGGTGAAC	CCGGACCCAC	TGGCCTGCCC	960
GGACCCCCTG	GCGAGCGTGG	TGGACCTGGT	AGCCGTGGTT	TCCCTGGCGC	AGATGGTGTT	1020
GCTGGTCCCA	AGGGTCCCGC	TGGTGAACGT	GGTTCTCCTG	GCCCCGCTGG	CCCCAAAGGA	1080
TCTCCTCGTG	AAGCTGGTCG	TCCCGGTGAA	GCTGGTCTGC	CTGGTGCCAA	GGGTCTGACT	1140
GGAAGCCCTG	GCAGCCCTGG	TCCTGATGGC	AAAACTGGCC	CCCCTGGTCC	CGCCGGTCAA	1200
GATGGTCGCC	CCGGACCCCC	AGGCCCACCT	GGTGCCCGTG	GTCAGGCTGG	TGTGATGGGA	1260
TTCCCTGGAC	CTAAAGGTGC	TGCTCGAGAG	CCCGGCAAGG	CTGGAGAGCG	AGGTGTTCCC	1320
GGACCCCCTC	GCGCTGTCGG	TCCTGCTGGC	AAAGATGGAG	AGGCTGGAGC	TCAGGGACCC	1380
CCTGGCCCTG	CTGGTCCCGC	TGGCGAGAGA	GGTGAACAAG	GCCCTGCTGG	CTCCCCCGGA	1440
TTCCAGGGTC	TCCCTGGTCC	TGCTGGTCCT	CCAGGTGAAG	CAGGCAAACC	TGGTGAACAG	1500
GGTGTTCCTG	GAGACCTTGG	CGCCCCTGGC	CCCTCTGGAG	CAAGAGGCGA	GAGAGGTTTC	1560
CCTGGCGAGC	GTGGTGTGCA	AGGTCCCCCT	GGTCCTGCTG	GACCCCGAGG	GGCCAACGGT	1620

GCTCCCGCCA	ACGATGCTGC	TAAGGGTGAT	GCTGGTGCCC	CTGGAGCTCC	CGGTAGCCAG	1680	
GGCGCCCCTG	GCCTTCAGGG	AATGCCTGGT	GAACGTGGŢG	CAGCTGGTCT	TCCAGGGCCT	1740	
AAGGGTGACA	GAGGTGATGC	TGGTCCCAAA	GGTGCTGATG	GCTCTCCTGG	CAAAGATGGC	1800	
GTCCGTGGTC	TGACCGACCC	CATTGGTCCT	CCTGGCCCTG	CTGGTGCCCC	TGGTGACAAG	1860	
GGTGAAAGTG	GTCCCAGCGG	CCCTGCTGGT	CCCACTGGAG	CTCGTGGTGC	CCCCGGAGAC	1920	
CGTGGTGAGC	CTGGTCCCCC	CGGCCCTGCT	GGCTTTGCTG	GCCCCCTGG	TGCTGACGGC	1980	
CAACCTGGTG	CTAAAGGCGA	ACCTGGTGAT	GCTGGTGCCA	AAGGCGATGC	TGGTCCCCCT	2040	
GGGCCTGCCG	GACCCGCTGG	ACCCCCTGGC	CCCATTGGTA	ATGTTGGTGC	TCCTGGAGCC	2100	
AAACGTGCTC	GCGGCAGCGC	TGGTCCCCCT	GGTGCTACTG	GTTTCCCTGG	TGCTGCTGGC	2160	
CGAGTCGGTC	CTCCTGGCCC	CTCTGGAAAT	GCTGGACCCC	CTGGCCCTCC	TGGTCCTGCT	2220	
GGCAAAGAAG	GCGGCAAAGG	TCCCCGTGGT	GAGACTGGCC	CTGCTGGACG	TCCTGGTGAA	2280	
GTTGGTCCCC	CTGGTCCCCC	TGGCCCTGCT	GGCGAGAAAG	GATCCCCTGG	TGCTGATGGT	2340	
CCTGCTGGTG	CTCCTGGTAC	TCCCGGGCCT	CAAGGTATTG	CTGGACAGCG	TGGTGTGGTC	2400	
GGCCTGCCTG	GTCAGAGAGG	AGAGAGAGGC	TTCCCTGGTC	TTCTTGGCCC	CTCTGGTGAA	2460	
CCTGGCAAAC	AAGGTCCCTC	TGGAGCAAGT	GGTGAACGTG	GTCCCCCGG	TCCCATGGGC	2520	
CCCCTGGAT	TGGCTGGACC	CCCTGGTGAA	TCTGGACGTG	AGGGGGCTCC	TGCTGCCGAA	2580	
GGTTCCCCTG	GACGAGACGG	TTCTCCTGGC	GCCAAGGGTG	ACCGTGGTGA	GACCGGCCCC	2640	

÷ 5

:10

. 

<u>్స్</u>

GCTGGACCCC	CTGGTGCTCC	TGGTGCTCCT	GGTGCCCCTG	GCCCCGTTGG	CCCTGCTGGC	2700
AAGAGTGGTG	ATCGTGGTGA	GACTGGTCCT	естестссе	CCGGTCCCGT	CGGCCCCGCT	2760
GGCGCCCGTG	GCCCGCCGG	ACCCCAAGGC	CCCCGTGGTG	ACAAGGGTGA	GACAGGCGAA	2820
CAGGGCGACA	GAGGCATAAA	GGGTCACCGT	GGCTTCTCTG	GCCTCCAGGG	TCCCCCTGGC	2880
CCTCCTGGCT	CTCCTGGTGA	ACAAGGTCCC	TCTGGAGCCT	CTGGTCCTGC	TGGTCCCCGA	2940
GGTCCCCCTG	GCTCTGCTGG	TGCTCCTGGC	AAAGATGGAC	TCAACGGTCT	CCCTGGCCCC	3000
ATTGGGCCCC	CTGGTCCTCG	CGGTCGCACT	GGTGATGCTG	GTCCTGTTGG	TCCCCCGGC	3-060
CCTCCTGGAC	CTCCTGGTCC	CCCTGGTCCT	CCCAGCGCTG	GTTTCGACTT	CAGCTTCCTC	3120
CCCCAGCCAC	CTCAAGAGAA	GGCTCACGAT	GGTGGCCGCT	ACTACCGGGC	TAGATCCGAT	3180
GAGGCTTCTG	GGATAGCCCC	AGAAGTTCCT	GATGACCGCG	ACTTCGAGCC	CTCCCTAGGC	3240
CCAGTGTGCC	CCTTCCGCTG	TCAATGCCAT	CTTCGAGTGG	TCCAGTGTTC	TGATTTGGGT	3300
CTGGACAAAG	TGCCAAAGGA	TCTTCCCCCT	GACACAACTC	TGCTAGACCT	GCAAAACAAC	3360
AAAATAACCG	AAATCAAAGA	TGGAGACTTT	AAGAACCTGA	AGAACCTTCA	CGCATTGATT	3420
CTTGTCAACA	ATAAAATTAG	CAAAGTTAGT	CCTGGAGCAT	TTACACCTTT	GGTGAAGTTG	3480
GAACGACTTT	ATCTGTCCAA	GAATCAGCTG	AAGGAATTGC	CAGAAAAAAT	GCCCAAAACT	3540
CTTCAGGAGC	TGCGTGCCCA	TGAGAATGAG	ATCACCAAAG	TGCGAAAAGT	TACTTTCAAT	3600
GGACTGAACC	AGATGATTGT	CATAGAACTG	GGCACCAATC	CGCTGAAGAG	CTCAGGAATT	3660

5	GAAAATGGGG CTTTCCAGGG AATGAAGAAG CTCTCCTACA TCCGCATTGC TGATACCAAT	3720
	ATCACCAGCA TTCCTCAAGG TCTTCCTCCT TCCCTTACGG AATTACATCT TGATGGCAAC	3780
10	AAAATCAGCA GAGTTGATGC AGCTAGCCTG AAAGGACTGA ATAATTTGGC TAAGTTGGGA	3840
	TTGAGTTTCA ACAGCATCTC TGCTGTTGAC AATGGCTCTC TGGCCAACAC GCCTCATCTG	3900
15 <sup>-</sup>	AGGGAGCTTC ACTTGGACAA CAACAAGCTT ACCAGAGTAC CTGGTGGGCT GGCAGAGCAT	3960
20	AAGTACATCC AGGTTGTCTA CCTTCATAAC AACAATATCT CTGTAGTTGG ATCAAGTGAC	4020
	TTCTGCCCAC CTGGACACAA CACCAAAAAG GCTTCTTATT CGGGTGTGAG TCTTTTCAGC	4080
25 <sup>-</sup>	AACCCGGTCC AGTACTGGGA GATACAGCCA TCCACCTTCA GATGTGTCTA CGTGCGCTCT	4140
	GCCATTCAAC TCGGAAACTA TAAGTAA	4167
<b>30</b>	(2) INFORMATION FOR SEQ ID NO:13:	
35 <sup>.</sup>	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 3349 base pairs	
•	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
40	(D) TOPOLOGY: linear	
•	(ii) MOLECULE TYPE: cDNA	
45 <sup>-</sup>		
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:	
55 <sup>°</sup>	GGGAAGGATT TCCATTTCCC AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT	60

50·

55<sup>-</sup>

TTCCGTGCCT	GGCCCCATGG	GTCCCTCTGG	TCCTCGTGGT	CTCCCTGGCC	CCCCTGGTGC	120
ACCTGGTCCC	CAAGGCTTCC	AAGGTCCCCC	TGGTGAGCCT	GGCGAGCCTG	GAGCTTCAGG	180
TCCCATGGGT	CCCCGAGGTC	CCCCAGGTCC	CCCTGGAAAG	AATGGAGATG	ATGGGGAAGC	240
TGGAAAACCT	GGTCGTCCTG	GTGAGCGTGG	GCCTCCTGGG	CCTCAGGGTG	CTCGAGGATT	300
GCCCGGAACA	GCTGGCCTCC	CTGGAATGAA	GGGACACAGA	GGTTTCAGTG	GTTTGGATGG	360
TGCCAAGGGA	GATGCTGGTC	CTGCTGGTCC	TAAGGGTGAG	CCTGGCAGCC	CTGGTGAAAA	420
TGGAGCTCCT	GGTCAGATGG	GCCCCGTGG	CCTGCCTGGT	GAGAGAGGTC	GCCCTGGAGC	480
CCCTGGCCCT	GCTGGTGCTC	GTGGAAATGA	TGGTGCTACT	GGTGCTGCCG	GGCCCCCTGG	540
TCCCACCGGC	CCCGCTGGTC	CTCCTGGCTT	CCCTGGTGCT	GTTGGTGCTA	AGGGTGAAGC	600
TGGTCCCCAA	GGGCCCCGAG	GCTCTGAAGG	TCCCCAGGGT	GTGCGTGGTG	AGCCTGGCCC	660
CCCTGGCCCT	GCTGGTGCTG	CTGGCCCTGC	TGGAAACCCT	GGTGCTGATG	GACAGCCTGG	720
TGCTAAAGGT	GCCAATGGTG	CTCCTGGTAT	TGCTGGTGCT	CCTGGCTTCC	CTGGTGCCCG	780
AGGCCCCTCT	GGACCCCAGG	GCCCCGGCGG	CCCTCCTGGT	CCCAAGGGTA	ACAGCGGTGA	840
ACCTGGTGCT	CCTGGCAGCA	AAGGAGACAC	TGGTGCTAAG	GGAGAGCCTG	GCCCTGTTGG	900
TGTTCAAGGA	CCCCTGGCC	CTGCTGGAGA	GGAAGGAAAG	CGAGGAGCTC	GAGGTGAACC	960
CGGACCCACT	GGCCTGCCCG	GACCCCCTGG	CGAGCGTGGT	GGACCTGGTA	GCCGTGGTTT	1020
CCCTGGCGCA	GATGGTGTTG	CTGGTCCCAA	GGGTCCCGCT	GGTGAACGTG	GTTCTCCTGG	1080

↓. 

CCCCGC	TGGC	CCCAAAGGAT	CTCCTGGTGA	AGCTGGTCGT	CCCGGTGAAG	CTGGTCTGCC	1140
TGGTGC	CAAG	GGTCTGACTG	GAAGCCCTGG	CAGCCCTGGT	CCTGATGGCA	AAACTGGCCC	1200
CCCTGG	TCCC	GCCGGTCAAG	ATGGTCGCCC	CGGACCCCCA	GGCCCACCTG	GTGCCCGTGG	1260
TCAGGC	TGGT	GTGATGGGAT	TCCCTGGACC	TAAAGGTGCT	GCTGGAGAGC	CCGGCAAGGC	1320
TGGAGA	AGCGA	GGTGTTCCCG	GACCCCCTGG	CGCTGTCGGT	CCTGCTGGCA	AAGATGGAGA	1380
GGCTGG	SAGCT	CAGGGACCCC	CTGGCCCTGC	TGGTCCCGCT	GGCGAGAGAG	GTGAACAAGG	1440
CCCTGC	CTGGC	TCCCCCGGAT	TCCAGGGTCT	CCCTGGTCCT	GCTGGTCCTC	CAGGTGAAGC	1500
AGGCAZ	AACCT	GGTGAACAGG	GTGTTCCTGG	AGACCTTGGC	GCCCCTGGCC	CCTCTGGAGC	1560
AAGAGG	SCGAG	AGAGGTTTCC	CTGGCGAGCG	TGGTGTGCAA	GGTCCCCCTG	GTCCTGCTGG	1620
ACCCC	GAGGG	GCCAACGGTG	CTCCCGGCAA	CGATGGTGCT	AAGGGTGATG	CTGGTGCCCC	1680
TGGAG	CTCCC	GGTAGCCAGG	GCGCCCCTGG	CCTTCAGGGA	ATGCCTGGTG	AACGTGGTGC	1740
AGCTG	<b>GTCTT</b>	CCAGGGCCTA	AGGGTGACAG	AGGTGATGCT	GGTCCCAAAG	GTGCTGATGG	1800
CTCTC	CTGGC	AAAGATGGCG	TCCGTGGTCT	GACCGGCCCC	ATTGGTCCTC	CTGGCCCTGC	1860
TGGTG	CCCCT	GGTGACAAGG	GTGAAAGTGG	TCCCAGCGGC	CCTGCTGGTC	CCACTGGAGC	1920
TCGTG	GTGCC	CCCGGAGACC	GTGGTGAGCC	TGGTCCCCCC	GGCCCTGCTG	GCTTTGCTGG	1980
ccccc	CTGGT	GCTGACGGCC	AACCTGGTGC	TAAAGGCGAA	CCTGGTGATG	CTGGTGCCAA	2040
AGGCG.	ATGCT	GGTCCCCCTG	GGCCTGCCGG	ACCCGCTGGA	CCCCCTGGCC	CCATTGGTAA	2100

:. 

TGTTGGTGCT	CCTGGAGCCA	AAGGTGCTCG	CGGCAGCGCT	GGTCCCCCTG	GTGCTACTGG	2160
TTTCCCTGGT	GCTGCTGGCC	GAGTCGGTCC	TCCTGGCCÇC	TCTGGAAATG	CTGGACCCC	2220
TGGCCCTCCT	GGTCCTGCTG	GCAAAGAAGG	CGGCAAAGGT	CCCCGTGGTG	AGACTGGCCC	2280
TGCTGGACGT	CCTGGTGAAG	TTGGTCCCCC	TGGTCCCCCT	GGCCCTGCTG	GCGAGAAAGG	2340
ATCCCCTGGT	GCTGATGGTC	CTGCTGGTGC	TCCTGGTACT	CCCGGGCCTC	AAGGTATTGC	2400
TGGACAGCGT	GGTGTGGTCG	GCCTGCCTGG	TCAGAGAGGA	GAGAGAGGCT	TCCCTGGTCT	2460
TCCTGGCCCC	TCTGGTGAAC	CTGGCAAACA	AGGTCCCTCT	GGAGCAAGTG	GTGAACGTGG	2520
TCCCCCCGGT	CCCATGGGCC	CCCCTGGATT	GGCTGGACCC	CCTGGTGAAT	CTGGACGTGA	2580
GGGGGCTCCT	GCTGCCGAAG	GTTCCCCTGG	ACGAGACGGT	TCTCCTGGCG	CCAAGGGTGA	2640
CCGTGGTGAG	ACCGGCCCCG	CTGGACCCCC	TGGTGCTCCT	GGTGCTCCTG	GTGCCCCTGG	2700
CCCCGTTGGC	CCTGCTGGCA	AGAGTGGTGA	TCGTGGTGAG	ACTGGTCCTG	CTGGTCCCGC	2760
CGGTCCCGTC	GGCCCCGCTG	GCGCCCGTGG	CCCCGCCGGA	CCCCAAGGCC	CCCGTGGTGA	2820
CAAGGGTGAG	ACAGGCGAAC	AGGGCGACAG	AGGCATAAAG	GGTCACCGTG	GCTTCTCTGG	2880
CCTCCAGGGT	CCCCCTGGCC	CTCCTGGCTC	TCCTGGTGAA	CAAGGTCCCT	CTGGAGCCTC	2940
TGGTCCTGCT	GGTCCCCGAG	GTCCCCCTGG	CTCTGCTGGT	GCTCCTGGCA	AAGATGGACT	3000
CAACGGTCTC	CCTGGCCCCA	TTGGGCCCCC	TGGTCCTCGC	GGTCGCACTG	GTGATGCTGG	3060
TCCTGTTGGT	ccccccccc	CTCCTGGACC	TCCTGGTCCC	CCTGGTCCTC	CCAGCGCTGG	3120

	TTTCGACTTC AGCTTCCTCC CCCAGCCACC TCAAGAGAAG GCTCACGATG GTGGCCGCTA	3180
5	CTACCGGGCT AGATCTCCAA AGGATCTTCC CCCTGACACA ACTCTGCTAG ACCTGCAAAA	3240
10	CAACAAAATA ACCGAAATCA AAGATGGAGA CTTTAAGAAC CTGAAGAACC TTCACGCATT	3300
	GATTCTTGTC AACAATAAAA TTAGCAAAGT TAGTCCTGGA TAACTGCAG	3349
<sub>(</sub> 15	(2) INFORMATION FOR SEQ ID NO:14:	
	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 57 base pairs	
	(B) TYPE: nucleic acid	
•	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
25		
e.	(ii) MOLECULE TYPE: cDNA	
30		
26	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:	
35		
	ATCGAGGGAA GGATTTCAGA ATTCGGATCC TCTAGAGTCG ACCTGCAGGC AAGCTTG	57
40	(2) INFORMATION FOR SEQ ID NO:15:	
	(i) SEQUENCE CHARACTERISTICS:	
<b>45</b>	(A) LENGTH: 3171 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
50		
•	(ii) MOLECULE TYPE: cDNA	

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

~ 50

: 55

	CAGCTGTCTT	ATGGCTATGA	TGAGAAATCA	ACCGGAGGAA	TTTCCGTGCC	TGGCCCCATG	60
10	GGTCCCTCTG	GTCCTCGTGG	TCTCCCTGGC	CCCCCTGGTG	CACCTGGTCC	CCAAGGCTTC	120
•	CAAGGTCCCC	CTGGTGAGCC	TGGCGAGCCT	GGAGCTTCAG	GTCCCATGGG	TCCCCGAGGT	180
15	CCCCCAGGTC	CCCCTGGAAA	GAATGGAGAT	GATGGGGAAG	CTGGAAAACC	TGGTCGTCCT	240
20	GGTGAGCGTG	GGCCTCCTGG	GCCTCAGGGT	GCTCGAGGAT	TGCCCGGAAC	AGCTGGCCTC	300
	CCTGGAATGA	AGGGACACAG	AGGTTTCAGT	GGTTTGGATG	GTGCCAAGGG	AGATGCTGGT	360
25	CCTGCTGGTC	CTAAGGGTGA	GCCTGGCAGC	CCTGGTGAAA	ATGGAGCTCC	TGGTCAGATG	420
	GGCCCCCGTG	GCCTGCCTGG	TGAGAGAGGT	CGCCCTGGAG	CCCCTGGCCC	TGCTGGTGCT	480
30	CGTGGAAATG	ATGGTGCTAC	TGGTGCTGCC	GGGCCCCCTG	GTCCCACCGG	CCCCGCTGGT	540
35	CCTCCTGGCT	TCCCTGGTGC	TGTTGGTGCT	AAGGGTGAAG	CTGGTCCCCA	AGGGCCCCGA	600
	GGCTCTGAAG	GTCCCCAGGG	TGTGCGTGGT	GAGCCTGGCC	CCCCTGGCCC	TGCTGGTGCT	660
40 .	GCTGGCCCTG	CTGGAAACCC	TGGTGCTGAT	GGACAGCCTG	GTGCTAAAGG	TGCCAATGGT	720
45	GCTCCTGGTA	TTGCTGGTGC	TCCTGGCTTC	CCTGGTGCCC	GAGGCCCCTC	TGGACCCCAG	780
45	GGCCCCGGCG	GCCCTCCTGG	TCCCAAGGGT	AACAGCGGTG	AACCTGGTGC	TCCTGGCAGC	840
~	AAAGGAGACA	CTGGTGCTAA	GGGAGAGCCT	GGCCCTGTTG	GTGTTCAAGG	ACCCCCTGGC	900

960	TGGCCTGCCC	CCGGACCCAC	CGAGGTGAAC	GCGAGGAGCT	AGGAAGGAAA	CCTGCTGGAG
1020	AGATGGTGTT	TCCCTGGCGC	AGCCGTGGTT	TGGACCTGGT	GCGAGCGTGG	GGACCCCCTG
1080	CCCCAAAGGA	GCCCCGCTGG	GGTTCTCCTG	TGGTGAACGT	AGGGTCCCGC	GCTGGTCCCA
1140	GGGTCTGACT	CTGGTGCCAA	GCTGGTCTGC	TCCCGGTGAA	AAGCTGGTCG	TCTCCTGGTG
1200	CGCCGGTCAA	CCCCTGGTCC	AAAACTGGCC	TCCTGATGGC	GCAGCCCTGG	GGAAGCCCTG
1260	TGTGATGGGA	GTCAGGCTGG	GGTGCCCGTG	AGGCCCACCT	CCGGACCCCC	GATGGTCGCC
1320	AGGTGTTCCC	CTGGAGAGCG	CCCGGCAAGG	TGCTGGAGAG	CTAAAGGTGC	TTCCCTGGAC
1380	TCAGGGACCC	AGGCTGGAGC	AAAGATGGAG	TCCTGCTGGC	GCGCTGTCGG	GGACCCCCTG
1440	CTCCCCGGA	GCCCTGCTGG	GGTGAACAAG	TGGCGAGAGA	CTGGTCCCGC	CCTGGCCCTG
1500	TGGTGAACAG	CAGGCAAACC	CCAGGTGAAG	TGCTGGTCCT	TCCCTGGTCC	TTCCAGGGTC
1560	GAGAGGTTTC	CAAGAGGCGA	CCCTCTGGAG	CGCCCCTGGC	GAGACCTTGG	GGTGTTCCTG
1620	GGCCAACGGT	GACCCCGAGG	GGTCCTGCTG	AGGTCCCCCT	GTGGTGTGCA	CCTGGCGAGC
1680	CGGTAGCCAG	CTGGAGCTCC	GCTGGTGCCC	TAAGGGTGAT	ACGATGGTGC	GCTCCCGGCA
1740	TCCAGGGCCT	CAGCTGGTCT	GAACGTGGTG	AATGCCTGGT	GCCTTCAGGG	GGCGCCCCTG
1800	CAAAGATGGC	GCTCTCCTGG	GGTGCTGATG	TGGTCCCAAA	GAGGTGATGC	AAGGGTGACA
1860	TGGTGACAAG	CTGGTGCCCC	CCTGGCCCTG	CATTGGTCCT	TGACCGGCCC	GTCCGTGGTC
1920	CCCCGGAGAC	CTCGTGGTGC	CCCACTGGAG	CCCTGCTGGT	GTCCCAGCGG	GGTGAAAGTG

CGTGGTGAGC	CTGGTCCCCC	CGGCCCTGCT	GGCTTTGCTG	GCCCCCTGG	TGCTGACGGC	1980
CAACCTGGTG	CTAAAGGCGA	ACCTGGTGAT	GCTGGTGCÇA	AAGGCGATGC	TGGTCCCCT	2040
GGGCCTGCCG	GACCCGCTGG	ACCCCCTGGC	CCCATTGGTA	ATGTTGGTGC	TCCTGGAGCC	2100
AAAGGTGCTC	GCGGCAGCGC	TGGTCCCCCT	GGTGCTACTG	GTTTCCCTGG	TGCTGCTGGC	2160
CGAGTCGGTC	CTCCTGGCCC	CTCTGGAAAT	GCTGGACCCC	CTGGCCCTCC	TGGTCCTGCT	2220
GGCAAAGAAG	GCGGCAAAGG	TCCCCGTGGT	GAGACTGGCC	CTGCTGGACG	TCCTGGTGAA	2280
GTTGGTCCCC	CTGGTCCCCC	TGGCCCTGCT	GGCGAGAAAG	GATCCCCTGG	TGCTGATGGT	2340
CCTGCTGGTG	CTCCTGGTAC	TCCCGGGCCT	CAAGGTATTG	CTGGACAGCG	TGGTGTGGTC	2400
GGCCTGCCTG	GTCAGAGAGG	AGAGAGAGGC	TTCCCTGGTC	TTCCTGGCCC	CTCTGGTGAA	2460
CCTGGCAAAC	AAGGTCCCTC	TGGAGCAAGT	GGTGAACGTG	GTCCCCCGG	TCCCATGGGC	2520
CCCCTGGAT	TGGCTGGACC	CCCTGGTGAA	TCTGGACGTG	AGGGGGCTCC	TGCTGCCGAA	2580
GGTTCCCCTG	GACGAGACGG	TTCTCCTGGC	GCCAAGGGTG	ACCGTGGTGA	GACCGGCCCC	2640
GCTGGACCCC	CTGGTGCTCC	TGGTGCTCCT	GGTGCCCCTG	GCCCCGTTGG	CCCTGCTGGC	2700
AAGAGTGGTG	ATCGTGGTGA	GACTGGTCCT	GCTGGTCCCG	CCGGTCCCGT	CGGCCCGCT	2760
GGCGCCCGTG	GCCCCGCCGG	ACCCCAAGGC	CCCCGTGGTG	ACAAGGGTGA	GACAGGCGAA	2820
CAGGGCGACA	GAGGCATAAA	GGGTCACCGT	GGCTTCTCTG	GCCTCCAGGG	TCCCCCTGGC	2880
CCTCCTGGCT	CTCCTGGTGA	ACAAGGTCCC	TCTGGAGCCT	CTGGTCCTGC	TGGTCCCCGA	2940

	GGTCCCCCTG GCTCTGCTGG TGCTCCTGGC AAAGATGGAC TCAACGGTCT CCCTGGCCCC	3000
5	ATTGGGCCCC CTGGTCCTCG CGGTCGCACT GGTGATGCTG GTCCTGTTGG TCCCCCCGGC	3060
5- 10	CCTCCTGGAC CTCCTGGTCC CCCTGGTCCT CCCAGCGCTG GTTTCGACTT CAGCTTCCTC	3120
	CCCCAGCCAC CTCAAGAGAA GGCTCACGAT GGTGGCCGCT ACTACCGGGC T	3171
15	(2) INFORMATION FOR SEQ ID NO:16:	
	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 1057 amino acids	
	(B) TYPE: amino acid	•
	(C) STRANDEDNESS: single (D) TOPOLOGY: unknown	
25	(b) Topologi: unknown	
	(ii) MOLECULE TYPE: peptide	
•	,,	
30		
Ş.		
35'	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:	
	Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val	
40	1 5 10 15	
	Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro	
	20 25 30	
45		
	Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly	
	35 40 45	
50		
•	Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro	
	50 55 60	
55		

 $\vec{\beta}$ 

5	Pro	Gly	Lys	Asn	Gly	Asp	Asp	Gly	Glu	Ala	Gly	ГÄв	Pro	Gly	Arg	Pro
	65					70					75					80
	Gly	Glu	Arg	Gly	Pro	Pro	Gly	Pro	Gln	Gly	Ala	Arg	Gly	Leu	Pro	Gly
10				-	85					90					95	
	mh	212	Gly	T.Au	Dro	G) v	Mot	Lva	Glv	His	Ara	Glv	Phe	Ser	Glv	T.011
15	THE	MIG	Gly		PLO	GLY	1466	בינם	105		****9	<b>-</b>	••	110	ULY	Deu
				100					105					110		
				•	<b>-1</b>	•	-1-	<b>~</b> 1			<b>~1</b>	D	¥	<b>01</b>	<b>~1</b>	
	Asp	GIŸ		гåв	GIÀ	Asp	Ата		Pro	Ala	GIA	PIO		GIĀ	GIU	Pro
20			115					120					125			
	Gly	Ser	Pro	Gly	Glu	Asn	Gly	Ala	Pro	Gly	Gln	Met	Gly	Pro	Arg	Gly
25		130					135					140				
	Leu	Pro	Gly	Glu	Arg	Gly	Arg	Pro	Gly	Ala	Pro	Gly	Pro	Ala	Gly	Ala
	145					150					155					160
30																
	Arg	Gly	Asn	Asp	Gly	Ala	Thr	Gly	Ala	Ala	Gly	Pro	Pro	Gly	Pro	Thr
;.					165					170					175	
35																
	Glv	Pro	Ala	Glv	Pro	Pro	Glv	Phe	Pro	Gly	Ala	Val	Glv	Ala	Lvs	Glv
	OL,			180			,		185	2			,	190	•	
,. 																
40	<b>61.</b> .	210	<i>α</i> 1	Dro	C1 5	GI v	Dro	7	C1	Ser	G] u	G] v	Dro	G] n	Gl v	77a 7
	GIU	ALA	-	PLO	GIII	GIY	PLO		GTÅ	SEL	Giu	GLY		GIII	GLY	Val
			195					200					205			
45·								_							_	_
	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Ala	Gly	Pro	Ala
		210					215					220				
50																
50	Gly	Asn	Pro	Gly	Ala	Asp	Gly	Gln	Pro	Gly	Ala	Lys	Gly	Ala	Asn	Gly
	225					230					235					240

5	Ala	Pro	Gly	Ile	Ala	Gly	Ala	Pro	Gly	Phe	Pro	Gly	Ala	Arg	Gly	Pro
					245					250					255	
	Ser	Gly	Pro	Gln	Gly	Pro	Gly	Gly	Pro	Pro	Gly	Pro	Lys	Gly	Asn	Ser
10		•		260	_			-	265					270		
	C1.	G] v	Dro	Gly	בומ	Dro	Glv	Sar	Taya	Glv	Δgn	Thr	Glv	λla	Tara	Glv
15	GIŞ	GIU	275	GLy	714		Gry	280	Lys	ur,	пор		285	724	n y o	GLY
,,			215					260					203			
	_			_	•					_	_	~3				
	Glu		GIY	Pro	val	GIY		GIn	GIA	Pro	Pro		PIO	Ala	GIY	Glu
20		290					295					300				
	Glu	Gly	Lys	Arg	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Pro	Thr	Gly	Leu	Pro
25	305					310					315					320
	Gly	Pro	Pro	Gly	Glu	Arg	Gly	Gly	Pro	Gly	Ser	Arg	Gly	Phe	Pro	Gly
					325					330					335	
30																
	Ala	Asp	Gly	Val	Ala	Gly	Pro	Lys	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Ser
				340					345					350		
35																
	Pro	Glv	Pro	Ala	Glv	Pro	Lvs	Glv	Ser	Pro	Gly	Glu	Ala	Gly	Arg	Pro
		•	355		-		•	360			-		365			
•																
40	G]v	Gl 11	Δla	Gly	Len	Pro	Glv	Δla	Taye	Glv	Len	Thr	Glv	Ser	Pro	Glv
	Gry		nzu	01,	Dou		375	74.4	270	027	204	380	0-7			
		370					313					360				
45	_	_	<b>~</b> 3		•	<b>~</b> 3	•	-1	<b>~</b> 3		<b>D</b>	<b>a</b> 1	<b>5</b>	• • • •	<b>03</b>	<b>43</b>
			GIY	Pro	Asp		Lys	Thr	GIĀ	Pro		GTÅ	Pro	ATS	GIY	
	385					390					395					400
50																
50	Asp	Gly	Arg	Pro	_	Pro	Pro	Gly	Pro		Gly	Ala	Arg	Gly		Ala
					405					410					415	

5	Gly	Val	Met	Gly	Phe	Pro	Gly	Pro	Lys	Gly	Ala	Ala	Gly	Glu	Pro	Gly
ŭ				420					425					430		
•																
	Lys	Ala	Gly	Glu	Arg	Gly	Val	Pro	Gly	Pro	Pro	Gly	Ala	Val	Gly	Pro
10	•		435		_	-		440	•				445		•	
,																
	21-	al.	Tve	Asp	alv	Glu.	פומ	GI <sub>1</sub>	בות	Gln.	Glv	Pro	Pro	G1v	Dro	71-
45	AIA		Lys	vob	Gry	GIU	455	GLY	ALG	0111	G.T.J	460		OLY	710	MIG
15		450					400					460				
•						_				<b>_</b>	_			_		_
	Gly	Pro	Ala	Gly	Glu		Gly	Glu	Gln	Gly		Ala	Gly	Ser	Pro	Gly
20	465					470					475					480
					•											
	Phe	Gln	Gly	Leu	Pro	Gly	Pro	Ala	Gly	Pro	Pro	GJA	Glu	Ala	Gly	Lys
25					485					490					495	
25																
4	Pro	Gly	Glu	Gln	Gly	Val	Pro	Gly	Asp	Leu	Gly	Ala	Pro	Gly	Pro	Ser
				500					505					510		
30																
S	Glv	Ala	Ara	Gly	Glu	Arg	Glv	Phe	Pro	Glv	Glu	Arq	Glv	Val	Gln	Glv
	1		515	2				520				•	525			
35·			313					320								
		D	~1··	Pro	210	~1··	Dwo	7	~1·	71.	λαν	Cl v	<b>71</b> 9	Pro	<i>α</i> 1	7.00
<i>t</i> ·	PIO		GIY	PIO	Ala	GIY		Arg	GIY	ALG	ABII		ALG	FIO	GIŞ	ASII
		530					535					540				
40					_		_	_	_							
	Asp	Gly	Ala	Lys	Gly	_	Ala	Gly	Ala	Pro	_	Ala	Pro	Gly	Ser	
	545					550					555					560
45																
	Gly	Ala	Pro	Gly	Leu	Gln	Gly	Met	Pro	Gly	Glu	Arg	Gly	Ala	Ala	Gly
					565					570					575	
•																
50	Leu	Pro	Gly	Pro	Lys	Gly	Asp	Arg	Gly	Asp	Ala	Gly	Pro	Lys	Gly	Ala
i				580					585		•			590		

5	Asp	Gly	Ser	Pro	Gly	Lys	Asp	Gly	Val	Arg	Gly	Leu	Thr	Gly	Pro	Ile
3			595					600					605			
•	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Pro	Gly	Asp	Lys	Gly	Glu	Ser	Gly
10		610					615					620				
;	Pro	Ser	Gly	Pro	Ala	Gly	Pro	Thr	Gly	Ala	Arg	Gly	Ala	Pro	Gly	Asp
15	625					630					635				•	640
	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Phe	Ala	Gly	Pro	Pro
••	_	-			645			•		650	•			•	655	
20															•••	
	Glv	Ala	Asp	Glv	Gln	Pro	Glv	Ala	Ive	Glv	Glu	Pro	Glv	Asn	Δla	Gly
	<b>4-7</b>			660			<b>-</b> -,		665				0-7	670		G <sub>1</sub> y
25				•••					003					070		
	בומ	Tara	Glv	Agn	212	Gly	Bro	Dro	G) v	Bro	212	Glv	Dro	Ala	C3	D
•	AIG	пåа	675	ADD.	Ald	GLY	210	680	GIY	PLO	ALG	GLY	685	ALA	GIY	PIO
30	•		6/3					000					003			
	D	<b>~</b> 1	D	T1 a	<b>~1</b>	3	**- *	<b>61</b>		<b>D</b>	<b>a</b> 1	<b>.</b> 1.	<b>T</b>	<b>~1</b>		_
	PIO		PIO	TTE	GIY	ABII		GIY	Ala	PIO	GIŞ		гув	Gly	Ala	Arg
		690					695					700				
35					_			_	_	_	_		_	_		
		Ser	Ala	Gly	Pro		Gly	Ala	Thr	Gly		Pro	Gly	Ala	Ala	Gly
•	705					710					715					720
40																
	Arg	Val	Gly	Pro	Pro	Gly	Pro	Ser	Gly	Asn	Ala	Gly	Pro	Pro	Gly	Pro
					725					730					735	
45																
	Pro	Gly	Pro	Ala	Gly	Lys	Glu	Gly	Gly	Lys	Gly	Pro	Arg	Gly	Glu	Thr
				740					745					750		
50	Gly	Pro	Ala	Gly	Arg	Pro	Gly	Glu	Val	Gly	Pro	Pro	Gly	Pro	Pro	Gly
			755					760					765			
•																
55																

. 5	Pro	Ala 770	Gly	Glu	Lys	Gly	Ser 775	Pro	Gly	Ala	Asp	Gly 780	Pro	Ala	Gly	Ala
	_		_,	_		_				•			_			
10	Pro 785	GIÀ	Thr	Pro	GIÀ	790	Gln	GIA	Ile	Ala	795	Gin	Arg	Gly	Val	Val 800
15	Gly	Leu	Pro	Gly	Gln 805	Arg	Gly	Glu	Arg	Gly 810	Phe	Pro	Gly	Leu	Pro 815	Gly
	Pro	Ser	Glv	Glu		Glv	Lva	Gln	Glv		Ser	Glv	Ala	Ser	Gly	Glu
20			,	820		,	,-		825			1		830	ory	314
25	Arg	Gly	Pro 835	Pro	Gly	Pro	Met	Gly 840	Pro	Pro	Gly	Leu	Ala 845	Gly	Pro	Pro
30	Gly	Glu 850	Ser	Gly	Arg	Glu	Gly 855	Ala	Pro	Ala	Ala	Glu 860	Gly	Ser	Pro	Gly
	Arg 865	Asp	Gly	Ser	Pro	Gly 870	Ala	Lys	Gly	Asp	Arg 875	Gly	Glu	Thr	Gly	Pro 880
35	Ala	Gly	Pro	Pro	Gly 885	Ala	Pro	Gly	Ala	Pro 890	Gly	Ala	Pro	Gly	Pro 895	Val
40	Gly	Pro	Ala	Gly 900	Lys	Ser	Gly	Ąsp	Arg 905	Gly	Glu	Thr	Gly	Pro 910	Ala	Gly
45	Pro	Ala	Gly		Val	Gly	Pro	Ala		Ala	Arg	Gly	Pro		Gly	Pro
50	<b>63</b>	<b>03</b>	915	•	, 	•		920		_,		_,	925			
	GIN	930 GIA	Pro	arg	θīλ	Asp	Lys 935	GIÀ	GIu	Thr	GIA	Glu 940	Gin	GIÀ	qaA	Arg

		Gly	Ile	Lys	Gly	His	Arg	Gly	Phe	ser	Gly	Leu	Gln	Gly	Pro	Pro	Gly	
<u>†</u> 5		945					950					955					960	
		Pro	Pro	Gly	Ser		Gly	Glu	Gln	Gly		Ser	Gly	Ala	Ser	Gly	Pro	
: 10						965					970					975		
		λla	Glv	Pro	Arg	Glv	Pro	Pro	Glv	Ser	Ala	Glv	Ala	Pro	Glv	Lys	Asp	
15		7,24	017	•••	980	1			,	985					990	-1-		
.0																		
		Gly	Leu	Asn	Gly	Leu	Pro	Gly	Pro	Ile	Gly	Pro	Pro	Gly	Pro	Arg	Gly	
20				995					100	0				100	5			
										_								
		Arg			Asp	Ala	Gly			Gly	Pro	Pro			Pro	Gly	Pro	
25			101	U				101	5				102	U				
		Pro	Gly	Pro	Pro	Gly	Pro	Pro	Ser	Ala	Gly	Phe	Asp	Phe	Ser	Phe	Leu	
,		102	-			-	103				_	103					1040	
30																		
		Pro	Gln	Pro	Pro	Gln	Glu	Lys	Ala	His	Asp	Gly	Gly	Arg	Tyr	Tyr	Arg	
						104	5				105	0				105	5	
35 <sup>,</sup>																		
		Ala																
40	(2)	INFO	RMAT	ION	FOR	SEQ	ID N	0:17	:									
45		(i)			E CH													
.0					ngth				cids									
			•-	•	PE: RAND				le.									
50					POLO			_	10									
			•-															
		(ii)	MOL	ECUL	E TY	PE:	pept	ide										

_	(1x) FEATURE:	
. 5	(A) NAME/KEY: Region	
	(B) LOCATION: 12	
	(D) OTHER INFORMATION: /note= "Amino acid sequence for	
10	glutathione S-transferase"	
	(ix) FEATURE:	
15	(A) NAME/KEY: Region	
	(B) LOCATION: 1920	
	(D) OTHER INFORMATION: /note= "338 repeats of the	
20	following triplet Gly-X-y wherein about 35% of the X and Y	
20	positions are occupied by proline and 4-hydroxyproline. "	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:	
	Xaa Met Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile	
30	1 5 10 15	
	Ser Val Pro Xaa Ser Ala Gly Phe Asp Phe Ser Phe Leu Pro Gln Pro	
35	20 25 30	
	Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg Ala	
40	35 40 45	
26		
	(2) INFORMATION FOR SEQ ID NO:18:	
45		
40	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 31 amino acids	
	(B) TYPE: amino acid	
50	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: unknown	
	(ii) MOI POUT P. TVDP. poptido	
55	(ii) MOLECULE TYPE: peptide	

	(ix) FEATURE:
5	(A) NAME/KEY: Region .
	(B) LOCATION: 12
	(D) OTHER INFORMATION: /note= "Amino acid sequence for
10	glutathione S-transferase."
1	
	(ix) FEATURE:
15	(A) NAME/KEY: Region
	(B) LOCATION: 45
	(D) OTHER INFORMATION: /note= "338 repeats of the
20	following triplet Gly-X-Y wherein about 35% of the X and Y
	positions are occupied by proline and 4-hydroxyproline. "
25	(with appropriate percentages) and the
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:
	Xaa Met Gly Xaa Tyr Ser Ala Gly Phe Asp Phe Ser Phe Leu Pro Gln
30	1 5 10 15
. '	Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg Ala
35	20 25 30
	(2) INFORMATION FOR SEQ ID NO:19:
40 <sup>.</sup>	·
	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 3171 base pairs
<b>4</b> 5	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear
50	(ii) MOLECULE TYPE: DNA (genomic)
-	(II) MODECODE TIPE: DNA (GENOMIC)

60

120

180

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

1 5

10

CAGCTGAGCT ATGGCTATGA TGAAAAAAGC ACCGGCGGCA TCAGCGTGCC GGGCCCGATG GGTCCGAGCG GCCCTCGTGG CCTGCCGGGC CCGCCAGGTG CGCCCGGTCC GCAGGGCTTT CAGGGTCCGC CGGGCGAACC GGGCGAACCT GGTGCGAGCG GCCCGATGGG CCCGCGCGGC CCGCCGGGTC CGCCAGGCAA AAACGGCGAT GATGGCGAAG CGGGCAAACC GGGACGTCCG GGTGAACGTG GCCCCCGGG CCCGCAGGGC GCGCGCGGAC TGCCGGGTAC TGCGGGACTG CCGGGCATGA AAGGCCACCG CGGTTTCTCT GGTCTGGATG GTGCGAAAGG TGATGCGGGT CCGGCGGGTC CGAAAGGTGA GCCGGGCAGC CCGGGCGAAA ACGGCGCGCC GGGTCAGATG

25

**35** ·

20

30

40

45

50

55

240 300 420 GGCCCGCGTG GCCTGCCTGG TGAACGCGGT CGCCCGGGCG CCCCGGGCCC AGCTGGCGCA 480 CGTGGCAACG ATGGTGCGAC CGGTGCGGCC GGTCCACCGG GCCCGACGGG CCCGGCGGGT 540 CCCCGGGCT TTCCGGGTGC GGTGGGTGCG AAAGGCGAAG CAGGTCCGCA GGGGCCGCGC 600 GGGAGCGAGG GTCCTCAGGG CGTTCGTGGT GAACCGGGCC CGCCGGGCCC GGCGGGTGCG 660 GCGGGCCCGG CTGGTAACCC TGGCGCGGAC GGTCAGCCAG GTGCGAAAGG TGCCAACGGC GCGCCGGGTA TTGCAGGTGC ACCGGGCTTC CCGGGTGCCC GCGGCCCGTC CGGCCCGCAG 780 GGCCCGGGCG GCCCGCCCGG CCCGAAAGGG AACAGCGGTG AACCGGGTGC GCCAGGCAGC 840 900 AAAGGCGACA CCGGTGCGAA AGGTGAACCG GGCCCAGTGG GTGTTCAAGG CCCGCCGGGC CCGGCGGGCG AGGAAGGCAA ACGCGGTGCT CGCGGTGAAC CGGGCCCGAC CGGCCTGCCT 960

 $f_{n} f_{n}$ 

40:

GGCCCGCCGG	GAGAACGTGG	TGGCCCGGGT	AGCCGCGGTT	TTCCGGGCGC	GGATGGTGTG	1020
GCGGGCCCGA	AAGGTCCGGC	GGGTGAACGT	GGTAGCCCĢG	GCCCGGCGGG	CCCAAAAGGC	1080
AGCCCGGGCG	AGGCAGGACG	TCCGGGTGAA	GCGGGTCTCC	CGGGCGCCAA	AGGTCTGACC	1140
GGCTCTCCGG	GCAGCCCGGG	TCCGGATGGC	AAAACGGGCC	CGCCTGGTCC	GGCCGGCCAG	1200
GATGGTCGCC	CGGGCCCGCC	GGGCCCGCCG	GGTGCCCGTG	GTCAGGCGGG	TGTCATGGGC	1260
TTTCCAGGCC	CCAAAGGTGC	GGCGGGTGAA	CCGGGCAAAG	CGGGCGAACG	CGGTGTCCCG	1320
GGTCCGCCGG	GCGCTGTCGG	GCCGGCGGGC	AAAGATGGCG	AAGCGGGCGC	GCAAGGCCCG	1380
CCGGGACCAG	CGGGTCCGGC	GGGCGAGCGC	GGTGAACAGG	GCCCGGCAGG	CAGCCCGGGT	1440
TTCCAGGGTC	TGCCGGGCCC	TGCGGGTCCA	CCGGGTGAAG	CGGGCAAACC	GGGGGAACAA	1500
GGTGTGCCGG	GCGACCTGGG	CGCCCCAGGC	CCGAGCGGCG	CGCGCGGCGA	ACGCGGTTTC	1560
CCGGGCGAAC	GTGGTGTGCA	GGGCCCGCCC	GGCCCGGCTG	GTCCGCGCGG	CGCCAACGGC	1620
GCGCCGGGCA	ACGATGGTGC	GAAAGGTGAT	GCGGGTGCCC	CAGGTGCGCC	GGGCAGCCAG	1680
GGCGCCCCGG	GGCTGCAAGG	CATGCCGGGT	GAACGTGGTG	CCGCGGGTCT	ACCGGGTCCG	1740
AAAGGCGACC	GCGGTGATGC	GGGTCCAAAA	GGTGCGGATG	GCTCCCCTGG	CAAAGATGGC	1800
GTTCGTGGTC	TGACCGGCCC	GATCGGCCCG	ccegecccee	CAGGTGCCCC	GGGTGACAAA	1860
GGTGAAAGCG	GTCCGAGCGG	CCCAGCGGGC	CCCACTGGTG	CGCGTGGTGC	CCCGGGCGAC	1920
CGTGGTGAAC	CGGGTCCGCC	GGGCCCGGCG	GGCTTTGCGG	GCCCGCCAGG	CGCTGACGGC	1980

ţ,

55

5	CAGCCGGGTG	CGAAAGGCGA	ACCGGGGGAT	GCGGGTGCTA	AAGGCGACGC	GGGTCCGCCG	2040
v.	GGCCCTGCCG	GCCCGGCGGG	CCCGCCAGGC	CCGATTGGCA	ACGTGGGTGC	GCCGGGTGCC	2100
10	AAAGGTGCGC	GCGGCAGCGC	TGGTCCGCCG	GGCGCGACCG	GTTTCCCCGG	TGCGGCGGG	2160
<u> 6</u>	CGCGTGGGTC	CGCCAGGCCC	GAGCGGTAAC	GCGGGTCCGC	CAGGTCCGCC	TGGCCCGGCT	2220
15	GGCAAAGAGG	GCGGCAAAGG	TCCGCGTGGT	GAAACCGGCC	CTGCGGGACG	TCCAGGTGAA	2280
20.	GTGGGTCCGC	CGGGCCGCC	GGGCCCGGCG	GGCGAAAAAG	GTAGCCCGGG	TGCGGATGGT	2340
	CCCGCCGGTG	CGCCAGGCAC	GCCGGGTCCG	CAAGGTATCG	CTGGCCAGCG	TGGTGTCGTC	2400
25	GGGCTGCCGG	GTCAGCGCGG	CGAACGCGGC	TTTCCGGGTC	TGCCGGGCCC	GAGCGGTGAG	2460
30	CCGGGCAAAC	AGGGTCCATC	TGGCGCGAGC	GGTGAACGTG	GCCCGCCGGG	TCCCATGGGC	2520
30.	CCGCCGGGTC	TGGCGGGCCC	TCCGGGTGAA	AGCGGTCGTG	AAGGCGCGCC	GGGTGCCGAA	2580
35	GGCAGCCCAG	GCCGCGACGG	TAGCCCGGGG	GCCAAAGGGG	ATCGTGGTGA	AACCGGCCCG	2640
.•	GCGGGCCCC	CGGGTGCACC	GGGCGCGCCG	GGTGCCCCAG	GCCCGGTGGG	CCCGGCGGGC	2700
40	AAAAGCGGTG	ATCGTGGTGA	GACCGGTCCG	GCGGGCCCGG	CCGGTCCGGT	GGGCCCAGCG	2760
45	GGCGCCCGTG	GCCCGGCCGG	TCCGCAGGGC	CCGCGGGGTG	ACAAAGGTGA	AACGGGCGAA	2820
	CAGGGCGACC	GTGGCATTAA	AGGCCACCGT	GGCTTCAGCG	GCCTGCAGGG	TCCACCGGGC	2880
50	CCGCCGGGCA	GTCCGGGTGA	ACAGGGTCCG	TCCGGAGCCA	GCGGGCCGGC	GGGCCCACGC	2940
	GGTCCGCCGG	GCAGCGCGGG	CGCGCCGGGC	AAAGACGGTC	TGAACGGTCT	GCCGGGCCCG	3000

.. 

. 

ATCGGCCCG	c ceeec	CCACG CGG	CCGCACC	GGTGATGC	GG GTCCG	GTGGG T	CCCCCGGG	C 3060
CCGCCGGGC	C CGCCAG	GCCC GCC	GGGACCG	CCGAGCGC	GG GTTTC	GACTT C	AGCTTCCT	G 3120
CCGCAGCCG	C CGCAGO	GAGAA AGO	CGCACGAC	GGCGGTCG	CT ACTAC	CGTGC G	3	3171
(2) INFOR	MATION I	FOR SEQ 1	ID NO:20	:				
(i)	SEQUENCI	E CHARACT	reristic	S:				
<b>,</b> _,	(A) LE	NGTH: 10	57 amino	acids				
	(B) TY	PE: amino	acid					
•	(C) ST	RANDEDNES	SS: sing	le				
	(D) TO	POLOGY: 1	ınknown					•
	·	E TYPE: ¡						
(xi)	SEQUENC	E DESCRI	PTION: S	EQ ID NO:	20:			
Gln	Leu Ser	Tyr Gly	Tyr Asp	Glu Lys	Ser Thr	Gly Gly	y Ile Ser	Val
1		5			10		15	
Pro	Gly Pro	Met Gly	Pro Ser	Gly Pro	Arg Gly	Leu Pro	o Gly Pro	Pro
Gly	Ala Pro	Gly Pro	Gln Gly	Phe Gln	Gly Pro	Pro Gl	y Glu Pro	Gly
, .	35		· ,	40		45		
Glu	Pro Gly	Ala Ser	Gly Pro	Met Gly	Pro Arg	Gly Pro	o Pro Gly	Pro
	50		55			60		

5	Pro	Gly	Lys	Asn	Gly	Asp	Asp	Gly	Glu	Ala	Gly	Lys	Pro	Gly	Arg	Pro
3	65					70					75					80
•	Gly	Glu	Arg	Gly	Pro	Pro	Gly	Pro	Gln	Gly	Ala	Arg	Gly	Leu	Pro	Glv
10	_		_	•	85		-			90			•		95	
										-					73	
· 	mb	21-	<b>a</b> 1	T 011	Dwa	a1	14a h	*	<b>~1</b>	***	B	<b>a</b> 1	nh -			_
	THE	Ala	GIĀ		PIQ	GIY	met	rya		uis	Arg	GIY	Pne	Ser	GIY	Leu
15				100					105					110		
,	Asp	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Glu	Pro
20			115					120					125			
	Gly	Ser	Pro	Gly	Glu	Asn	Gly	Ala	Pro	Gly	Gln	Met	Gly	Pro	Arg	Gly
		130					135					140				
25																
	Leu	Pro	Gly	Glu	Arg	Gly	Arg	Pro	Gly	Ala	Pro	Gly	Pro	Ala	Glv	Ala
;	145		_			150			-		155	•				160
30																100
	D.r.a	GIV	λen	Dan	Glv	בוג	Thr	G1v	77-	חות	C1.,	Dvo	Dwo	Gly		Min
	ΑLY	GIŢ	<b></b>	nop	165	ALG	1111	Gry	ALG		GIY	PIO	PLO	GLY		Thr
					102					170					175	
35					_		_									
	GIÅ	Pro	Ala		Pro	Pro	Gly	Phe	Pro	Gly	Ala	Val	Gly	Ala	Lys	Gly
<u> </u>				180					185					190		
40																
	Glu	Ala	Gly	Pro	Gln	Gly	Pro	Arg	Gly	Ser	Glu	Gly	Pro	Gln	Gly	Val
			195					200					205			
45	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Ala	Gly	Pro	Ala
		210					215				-	220				
50	Gly	Asn	Pro	Gly	Ala	Asp	Glv	Gln	Pro	Glv	Ala	Lvs	Glv	Ala	Asp	G] v
	225	-	-	•		230	,			1	235	-,-	,			240
											200					240

55

çi.

	Ala	Pro	Gly	Ile	Ala	Gly	Ala	Pro	Gly	Phe	Pro	Gly	Ala	Arg	Gly	Pro
5					245					250					255	
	ser	Gly	Pro	Gln	Gly	Pro	Gly	Gly	Pro	Pro	Gly	Pro	Lys	Gly	naA	Ser
10				260					265					270		
	Gly	Glu	Pro	Gly	Ala	Pro	Gly	Ser	Lys	Gly	Asp	Thr	Gly	Ala	Lys	Gly
15			275					280					285			
	Glu	Pro	Gly	Pro	Val	Gly	Val	Gln	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Glu
		290					295					300				
20																
٠.	Glu	Gly	Lys	Arg	Gly		Arg	Gly	Glu	Pro	Gly	Pro	Thr	Gly	Leu	Pro
25	305					310					315					320
	Gly	Pro	Pro	Gly	Glu	Arg	Gly	Gly	Pro	Gly	Ser	Arg	Gly	Phe	Pro	Gly
	-			_	325					330					335	
30																
	Ala	Asp	Gly	Val	Ala	Gly	Pro	Lys	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Ser
				340					345					350		
35	Pro	Glv	Pro	Δla	Glv	Pro	Lvs	Ġlv	Ser	Pro	Gly	Glu	Ala	Glv	Arg	Pro
:	110	UL,	355		U-,	120	_,_	360			,		365		3	
40	Gly	Glu	Ala	Gly	Leu	Pro	Gly	Ala	Lys	Gly	Leu	Thr	Gly	Ser	Pro	Gly
		370					375					380				
<b>4</b> 5	Ser	Pro	Glv	Pro	asA	Glv	Lvs	Thr	Glv	Pro	Pro	Glv	Pro	Ala	Glv	Gln
	385		017			390	2,5		,		395	,			1	400
50	Asp	Gly	Arg	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Ala	Arg	Gly	Gln	Ala
					405					410					415	

	Gly	Val	Met	Gly	Phe	Pro	Gly	Pro	Lys	Gly	Ala	Ala	Gly	Glu	Pro	Gly
5				420					425					430		
	Lys	Ala	Gly	Glu	Arg	Gly	Val	Pro	Gly	Pro	Pro	Gly	Ala	Val	Gly	Pro
10			435					440					445			
	Ala	Glv	Lys	Asp	Gly	Glu	Ala	Glv	Ala	Gln	Gly	Pro	Pro	Gly	Pro	Ala
		450	•	•	•		455				•	460		,		
15																
	Glv	Pro	Ala	Glv	Glu	Ara	Glv	Glu	Gln	Glv	Pro	Ala	Glv	Ser	Pro	G) v
	465			,		470	,			1	475					480
20	103										175					400
	Dho	Cln.	C1	T au	Dro	C1	Dwa	21-	<b>~</b> 1	Dwa	Dwa	~1··	<b>a</b> 1	Ala	<b>~</b> 1	•
	Pile	GIII	GIY	neu		GIY	PIQ	AId	GIY		PIO	GIY	GIU	ATA	•	ràs
25					485					490					495	
		_			_	_										
,	Pro	Gly	Glu		Gly	Val	Pro	Gly		Leu	Gly	Ala	Pro	Gly	Pro	Ser
.'				500					505					510		
30																
	Gly	Ala	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Glu	Arg	Gly	Val	Gln	Gly
			515					520					525			
35																
	Pro	Pro	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Ala	Asn	Gly	Ala	Pro	Gly	Asn
		530					535					540				
40	Asp	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ser	Gln
	545					550					555					560
45	Gly	Ala	Pro	Gly	Leu	Gln	Gly	Met	Pro	Gly	Glu	Arq	Glv	Ala	Ala	Glv
	•			_	565		-			570		-	•		575	•
50	Leu	Pro	G] v	Pro	Lva	G) v	Ago	Ara	Glv	Agn	Ala	Glv	Pro	Lys	Glv	Δla
50			,	580	-,-	1	.wp	y	585	p	-TA-U	,	0	590	J-7	A.a
				200					د د د					J J (		

5	Asp	Gly	Ser 595	Pro	Gly	Lys	Asp		Val	Arg	Gly	Leu		Gly	Pro	Ile
			373					600					605			
10	Gly		Pro	Gly	Pro	Ala		Ala	Pro	Gly	Asp		Gly	Glu	Ser	Gly
		610					615					620				
15		Ser	Gly	Pro	Ala		Pro	Thr	Gly	Ala		Gly	Ala	Pro	Gly	_
	625					630					635					640
	Arg	Gly	Glu	Pro		Pro	Pro	Gly	Pro		Gly	Phe	Ala	Gly	Pro	Pro
20					645					650					655	
	Gly	Ala	Asp		Gln	Pro	Gly	Ala		Gly	Glu	Pro	Gly		Ala	Gly
25				660					665					670		
	Ala	Lys	Gly	Asp	Ala	Gly	Pro		Gly	Pro	Ala	Gly		Ala	Gly	Pro
30			675					680					685			
	Pro		Pro	Ile	Gly	Asn		Gly	Ala	Pro	Gly		Lys	Gly	Ala	Arg
35		690					695					700				
:•		Ser	Ala	Gly	Pro		Gly	Ala	Thr	Gly		Pro	Gly	Ala	Ala	
40	705					710					715					720
	Arg	Val	Gly	Pro		Gly	Pro	Ser	Gly		Ala	Gly	Pro	Pro	•	Pro
45					725					730					735	
	Pro	Gly	Pro	•	Gly	Lys	Glu	Gly		Lys	Gly	Pro	Arg		Glu	Thr
				740					745					750		
50	Gly	Pro	Ala	Gly	Arg	Pro	Gly		Val	Gly	Pro	Pro		Pro	Pro	Gly
			755					760					765			

5	Pro	Ala	Gly	Glu	Lys	Gly	Ser	Pro	Gly	Ala	Asp	Gly	Pro	Ala	Gly	Ala
		770					775					780				
10	Pro	Gly	Thr	Pro	Gly	Pro	Gln	Gly	Ile	Ala	Gly	Gln	Arg	Gly	Val	Val
,,,	785					790					795					800
	Gly	Leu	Pro	Gly	Gln	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Leu	Pro	Gly
15					805					810					815	
	Pro	Ser	Gly	Glu	Pro	Gly	Lys	Gln	Gly	Pro	Ser	Gly	Ala	Ser	Gly	Glu
20				820					825					830		
	Arg	Gly	Pro	Pro	Gly	Pro	Met	Gly	Pro	Pro	Gly	Leu	Ala	Gly	Pro	Pro
25			835					840					845			
	Gly	Glu	Ser	Gly	Arg	Glu	Gly	Ala	Pro	Gly	Ala	Glu	Gly	Ser	Pro	Gly
30		850					855					860				
	Arg	Asp	Gly	Ser	Pro	Gly	Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro
35	865					870					875					880
	Ala	Gly	Pro	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Pro	Val
;. 40					885					890					895	
••	Gly	Pro	Ala	Gly	Lys	Ser	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro	Ala	Gly
				900					905					910		
45 ·	Pro	Ala	Gly	Pro	Val	Gly	Pro	Ala	Gly	Ala	Arg	Gly	Pro	Ala	Gly	Pro
			915					920					925			
50	Gln	Gly	Pro	Arg	Gly	Asp	Lys	Gly	Glu	Thr	Gly	Glu	Gln	Gly	Asp	Arg
		930					935					940				

.í . <b>5</b>		Gly 945	Ile	Lys	Gly	His	Arg 950	Gly	Phe	Ser	Gly	Leu 955	Gln	Gly	Pro	Pro	Gly 960
10		Pro	Pro	Gly	Ser	Pro 965	Gly	Glu	Gln	Gly	Pro 970	Ser	Gly	Ala	Ser	Gly 975	Pro
15		Ala	Gly	Pro	Arg 980	Gly	Pro	Pro	Gly	Ser 985	Ala	Gly	Ala	Pro	Gly 990	Lys	Asp
20		Gly	Leu	Asn 995	Gly	Leu	Pro	Gly	Pro 1000		Gly	Pro	Pro	Gly 1009		Arg	Gly
25		Arg	Thr 1010		Ąsp	Ala	Gly	Pro 1019		Gly	Pro	Pro	Gly 1020	Pro	Pro	Gly	Pro
30		Pro 1025	-	Pro	Pro	Gly	∤ro 1030		Ser	Ala	Gly	Phe 1035	_	Phe	Ser	Phe	Leu 1040
-: 35		Pro	Gln	Pro	Pro	Gln 1049		Lys	Ala	His	Asp 1050		Gly	Arg	Tyr	Tyr 105	-
40		Ala			-												
	(2)	INFO			FOR S												
45			(B)	TY	NGTH PE: 1 RANDI	nucle	eic a	cid									
		(ii)	(D)	TO	POLO	GY: ]	linea	_									

4

55.

<b>5</b> .	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:	
	GGAATTCATG CAGCTGAGCT ATGGCTATGA TGAAAAAAGC ACCGGCGGCA TCAGCGTGCC	60
,,	GGGCCCGATG GGTCCGAGC	79
15.	(2) INFORMATION FOR SEQ ID NO:22:	
20	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 75 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	·
25	(ii) MOLECULE TYPE: cDNA	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:	
· · · · · · · · · · · · · · · · · · ·	GGCCCGGGCT ACCCAGGCTC GCCGGGCGCA CCGGACGGCC CGGGCGGTCC AGCGGGGCCA	60
40	GCATTATTCG AACCC	75
ė.	(2) INFORMATION FOR SEQ ID NO:23:	
<b>45</b> . <b>50</b>	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 81 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	

5	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:	
10	GGAATTCCGG GTCCGCAGGG CTTTCAGGGT CCGCCGGGCG AACCTGGTGC GAGCGGCCCG	60
••	ATGGGCCCGC GCGGCCCGCC C	81
	(2) INFORMATION FOR SEQ ID NO:24:	
	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 87 base pairs	
	(B) TYPE: nucleic acid	•
	(C) STRANDEDNESS: single	
25	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
30		
:		
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:	
t,	TACCCGGGCG CGCCGGCGG CCCAGGCGGT CCGTTTTTGC CGCTACTACC GTTCGCCCGT	60
40	TTGGCCCTGC AGGCATTATT CGAACCC	87
•	(2) INFORMATION FOR SEQ ID NO:25:	
45	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 111 base pairs	
	(B) TYPE: nucleic acid	
50·	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	

5	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:	
10	CAGCTGAGCT ATGGCTATGA TGAAAAAAGC ACCGGCGGCA TCAGCGTGCC GGGCCCGATG	60
15	GGTCCGAGCG GCCCTCGTGG CCTGCCGGGC CCGCCAGGTG CGCCCGGTCC G	111
•	(2) INFORMATION FOR SEQ ID NO:26:	
20	<ul><li>(i) SEQUENCE CHARACTERISTICS:</li><li>(A) LENGTH: 37 amino acids</li><li>(B) TYPE: amino acid</li></ul>	
25	(C) STRANDEDNESS: single (D) TOPOLOGY: unknown	
30	(ii) MOLECULE TYPE: peptide	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:	
35	Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val 1 5 10 15	
40	Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro 20 25 30	
45	Gly Ala Pro Gly Pro	
	(2) INFORMATION FOR SEQ ID NO:27:	
50	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 240 base pairs  (B) TYPE: nucleic acid	
55	(0)	

5	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
2	<u>.                                    </u>	
10	(ii) MOLECULE TYPE: cDNA	
j.	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:	
15	CAGCTGAGCT ATGGCTATGA TGAAAAAAGC ACCGGCGGCA TCAGCGTGCC GGGCCCGATG	60
20	GGTCCGAGCG GCCCTCGTGG CCTGCCGGGC CCGCCAGGTG CGCCCGGTCC GCAGGGCTTT	120
	CAGGGTCCGC CGGGCGAACCT GGTGCGAGCG GCCCGATGGG CCCGCGCGCG	180
25	CCGCCGGGTC CGCCAGGCAA AAACGGCGAT GATGGCGAAG CGGGCAAACC GGGACGTCCG	240
w		
30	(2) INFORMATION FOR SEQ ID NO:28:	
•	(i) SEQUENCE CHARACTERISTICS:	
**	(A) LENGTH: 80 amino acids	
35		
,	(B) TYPE: amino acid	
₹.	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: unknown	
40		
	(ii) MOLECULE TYPE: peptide	
45	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:	
	Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val	
50	1 5 10 15	
	Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro	
	20 25 30	
55		

5	Gly	Ala	Pro	Gly	Pro	Gln	Gly	Phe	Gln	Gly	Pro	Pro	Gly	Glu	Pro	Gly	
J			35					40					45				
٠.	Glu	Pro	Gly	Ala	Ser	Gly	Pro	Met	Gly	Pro	Arg	Gly	Pro	Pro	Gly	Pro	
10		50					55					60					
.1	Pro	Gly	Lys	Asn	Gly	Asp	Asp	Gly	Glu	Ala	Gly	Lys	Pro	Gly	Arg	Pro	
15	65					70					75					80	
20	(2) INFO	RMAT:	ION 1	FOR S	SEQ :	ID NO	):29:	:									
•	(i)	SEQU	JENC!	E CHI	ARAC:	rer i s	STICS	S :									
25		(B)	TYI	ngth: PE: 1 RANDI POLOC	oucle EDNES	eic a	acid		3								
30	(ii)	MOLI	ECULI	E TYI	PE: (	:DNA											
35	(xi)	SEQU	UENCI	E DES	CRI	PTIO	1: SI	EQ II	ONO:	29:							
;	CAGTATGA'	TG G	AAAA	GGAG7	TG(	SACT	rggc	CCT	GACC	AA 7	rggg(	CTTA	AT GO	GAC	CTAGA		60
40	GGCCCACC'	TG G7	rgca(	GCTG(	G AGO	cccı	\GGC	CCT	CAAGG	er i	CCAJ	AGGA	CC TO	GCTG(	STGAG	}	120
45	CCTGGTGA	AC C	rggt(	CAAAC	TGO	STCCI	rgca	GGT	CTC	etg (	FTCC/	AGCTO	G C	CCTC	CTGGC	:	180
	AAGGCTGG	TG AJ	AGATY	GTC/	CCC	CTGG!	<b>LAAA</b>	ccc	GACG	BAC (	CTGG	(DAD	AG AC	3GAG1	TGT1	?	240
50	GGACCACA	GG G1	rgct(	CGTGC	FTT	rccci	rgga	ACTO	CTGG	SAC 7	rtcci	rggc	TT C	\AAG(	CATI	:	300
	AGGGGACA	CA AT	rggt(	CTGG	A TGC	SATTO	BAAG	GGA	CAGCO	CG (	TGC	CCTC	G TO	STGA	AGGGI	•	360
55																	

5	GAACCTGGTG	CCCCTGGTGA	AAATGGAACT	CCAGGTCAAA	CAGGAGCCCG	TGGGCTTCCT	420
	GGTGAGAGAG	GACGTGTTGG	TGCCCCTGGC	CCAGCTGGTG	CCCGTGGCAG	TGATGGAAGT	480
10	GTGGGTCCCG	TGGGTCCTGC	TGGTCCCATT	GGGTCTGCTG	GCCCTCCAGG	CTTCCCAGGT	540
15	GCCCCTGGCC	CCAAGGGTGA	AATTGGAGCT	GTTGGTAACG	CTGGTCCTGC	TGGTCCCGCC	600
	GGTCCCCGTG	GTGAAGTGGG	TCTTCCAGGC	CTCTCCGGCC	CCGTTGGACC	TCCTGGTAAT	660
20	CCTGGAGCAA	ACGGCCTTAC	TGGTGCCAAG	GGTGCTGCTG	GCCTTCCCGG	CGTTGCTGGG	720
	GCTCCCGGCC	TCCCTGGACC	CCGCGGTATT	CCTGGCCCTG	TTGGTGCTGC	CGGTGCTACT	780
<b>25</b>	GGTGCCAGAG	GACTTGTTGG	TGAGCCTGGT	CCAGCTGGCT	CCAAAGGAGA	GAGCGGTAAC	840
30	AAGGGTGAGC	CCGGCTCTGC	TGGGCCCCAA	GGTCCTCCTG	GTCCCAGTGG	TGAAGAAGGA	900
	AAGAGAGGCC	CTAATGGGGA	AGCTGGATCT	GCCGGCCCTC	CAGGACCTCC	TGGGCTGAGA	960
35 <sup>-</sup>	GGTAGTCCTG	GTTCTCGTGG	TCTTCCTGGA	GCTGATGGCA	GAGCTGGCGT	CATGGGCCCT	1020
	CCTGGTAGTC	GTGGTGCAAG	TGGCCCTGCT	GGAGTCCGAG	GACCTAATGG	AGATGCTGGT	1080
40	CGCCCTGGGG	AGCCTGGTCT	CATGGGACCC	AGAGGTCTTC	CTGGTTCCCC	TGGAAATATC	1140
45	GGCCCCGCTG	GAAAAGAAGG	TCCTGTCGGC	CTCCCTGGCA	TCGACGGCAG	GCCTGGCCCA	1200
	ATTGGCCCAG	CTGGAGCAAG	AGGAGAGCCT	GGCAACATTG	GATTCCCTGG	ACCCAAAGGC	1260
50	CCCACTGGTG	ATCCTGGCAA	AAACGGTGAT	AAAGGTCATG	CTGGTCTTGC	TGGTGCTCGG	1320
*	GGTGCTCCAG	GTCCTGATGG	AAACAATGGT	GCTCAGGGAC	CTCCTGGACC	ACAGGGTGTT	1380

CAAGGTGGAA	AAGGTGAACA	GGGTCCCGCT	GGTCCTCCAG	GCTTCCAGGG	TCTGCCTGGC	1440
CCCTCAGGTC	CCGCTGGTGA	AGTTGGCAAA	CCAGGAGAĄA	GGGGTCTCCA	TGGTGAGTTT	1500
GGTCTCCCTG	GTCCTGCTGG	TCCAAGAGGG	GAACGCGGTC	CCCCAGGTGA	GAGTGGTGCT	1560
GCCGGTCCTA	CTGGTCCTAT	TGGAAGCCGA	GGTCCTTCTG	GACCCCCAGG	GCCTGATGGA	1620
AACAAGGGTG	AACCTGGTGT	GGTTGGTGCT	GTGGGCACTG	CTGGTCCATC	TGGTCCTAGT	1680
GGACTCCCAG	GAGAGAGGG	TGCTGCTGGC	ATACCTGGAG	GCAAGGGAGA	AAAGGGTGAA	1740
CCTGGTCTCA	GAGGTGAAAT	TGGTAACCCT	GGCAGAGATG	GTGCTCGTGG	TGCTCATGGT	1800
GCTGTAGGTG	CCCCTGGTCC	TGCTGGAGCC	ACAGGTGACC	GGGGCGAAGC	TGGGGCTGCT	1860
GGTCCTGCTG	GTCCTGCTGG	TCCTCGGGGA	AGCCCTGGTG	AACGTGGCGA	GGTCGGTCCT	1920
GCTGGCCCCA	ACGGATTTGC	TGGTCCGGCT	GGTGCTGCTG	GTCAACCGGG	TGCTAAAGGA	1980
GAAAGAGGAG	CCAAAGGGCC	TAAGGGTGAA	AACGGTGTTG	TTGGTCCCAC	AGGCCCCGTT	2040
GGAGCTGCTG	GCCCAGCTGG	TCCAAATGGT	CCCCCGGTC	CTGCTGGAAG	TCGTGGTGAT	2100
GGAGGCCCCC	CTGGTATGAC	TGGTTTCCCT	GGTGCTGCTG	GACGGACTGG	TCCCCCAGGA	2160
CCCTCTGGTA	TTTCTGGCCC	TCCTGGTCCC	CCTGGTCCTG	CTGGGAAAGA	AGGGCTTCGT	2220
GGTCCTCGTG	GTGÀCCAAGG	TCCAGTTGGC	CGAACTGGAG	AAGTAGGTGC	AGTTGGTCCC	2280
CCTGGCTTCG	CTGGTGAGAA	GGGTCCCTCT	GGAGAGGCTG	GTACTGCTGG	ACCTCCTGGC	2340
ACTCCAGGTC	CTCAGGGTCT	TCTTGGTGCT	CCTGGTATTC	TGGGTCTCCC	TGGCTCGAGA	2400

GGTGAACGTG	GTCTACCTGG	TGTTGCTGGT	GCTGTGGGTG	AACCTGGTCC	TCTTGGCATT	2460
GCCGGCCCTC	CTGGGGCCCG	TGGTCCTCCT	GGTGCTGTGG	GTAGTCCTGG	AGTCAACGGT	2520
GCTCCTGGTG	AAGCTGGTCG	TGATGGCAAC	CCTGGGAACG	ATGGTCCCCC	AGGTCGCGAT	2580
GGTCAACCCG	GACACAAGGG	AGAGCGCGGT	TACCCTGGCA	ATATTGGTCC	CGTTGGTGCT	2640
GCAGGTGCAC	CTGGTCCTCA	TGGCCCCGTG	GGTCCTGCTG	GCAAACATGG	AAACCGTGGT	2700
GAAACTGGTC	CTTCTGGTCC	TGTTGGTCCT	GCTGGTGCTG	TTGGCCCAAG	AGGTCCTAGT	2760
GGCCCACAAG	GCATTCGTGG	CGATAAGGGA	GAGCCCGGTG	AAAAGGGGCC	CAGAGGTCTT	2820
CCTGGCTTAA	AGGGACACAA	TGGATTGCAA	GGTCTGCCTG	GTATCGCTGG	TCACCATGGT	2880
GATCAAGGTG	CTCCTGGCTC	CGTGGGTCCT	GCTGGTCCTA	GGGGCCCTGC	TGGTCCTTCT	2940
GGCCCTGCTG	GAAAAGATGG	TCGCACTGGA	CATCCTGGTA	CGGTTGGACC	TGCTGGCATT	3000
CGAGGCCCTC	AGGGTCACCA	AGGCCCTGCT	GCCCCCTG	GTCCCCCTGG	CCCTCCTGGA	3060
CCTCCAGGTG	TAAGCGGTGG	TGGTTATGAC	TTTGGTTACG	ATGGAGACTT	CTACAGGGCT	3120

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1040 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

55

10

5

10

15

20

25

30

35

45

<b>5</b>	(xi)	SEQ	JENCI	E DES	SCRII	PTIO	N: S	EQ II	ОИО	:30:						
.10	Gln 1	Tyr	Asp	Gly	Lys 5	Gly	Val	Gly	Leu	Gly 10	Pro	Gly	Pro	Met	Gly 15	Leu
	Met	Gly	Pro	Arg 20	Gly	Pro	Pro	Gly	Ala 25	Ala	Gly	Ala	Pro	Gly 30	Pro	Gln
·15	Gly	Phe	Gln 35	Gly	Pro	Ala	Gly	Glu 40	Pro	Gly	Glu	Pro	Gly 45	Gln	Thr	Gly
20	Pro	Ala 50	Gly	Ala	Arg	Gly	Pro 55	Ala	Gly	Pro	Pro	Gly 60	Lys	Ala	Gly	Glu
25	Asp 65	Gly	His	Pro	Gly	Lys 70	Pro	Gly	Arg	Pro	Gly 75	Glu	Arg	Gly	Val	Val 80
30	Gly	Pro	Gln	Gly	Ala 85	Arg	Gly	Phe	Pro	Gly 90	Thr	Pro	Gly	Leu	Pro 95	Gly
35	Phe	Lys	Gly	Ile 100	Arg	Gly	His	Asn	Gly 105	Leu	Asp	Gly	Leu	Lys 110	Gly	Gln
40	Pro	Gly	Ala 115	Pro	Gly	Val	Lys	Gly 120	Glu	Pro	Gly	Ala	Pro 125	Gly	Glu	Asn
45	Gly	Thr	Pro	Gly	Gln	Thr	Gly 135	Ala	Arg	Gly	Leu	Pro 140	Gly	Glu	Arg	Gly
50	Arg 145	Val	Gly	Ala	Pro	Gly 150	Pro	Ala	Gly	Ala	Arg 155	Gly	Ser	Asp	Glγ	Ser 160

	Val	Gly	Pro	Val	Gly	Pro	Ala	Gly	Pro	Ile	Gly	Ser	Ala	Gly	Pro	Pro
5					165					170					175	
<b>.</b>	Gly	Phe	Pro	Gly	Ala	Pro	Gly	Pro	Lys	Gly	Glu	Ile	Gly	Ala	Val	Gly
10				180					185					190		
	Asn	Ala	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Glu	Val	Gly	Leu
15			195					200					205			
	Pro		Leu	Ser	Gly	Pro		Gly	Pro	Pro	Gly	Asn	Pro	Gly	Ala	Asn
20		210					215					220				
		Leu	Thr	Gly	Ala		Gly	Ala	Ala	Gly		Pro	Gly	Val	Ala	Gly
25	225					230					235					240
	Ala	Pro	Gly	Leu		Gly	Pro	Arg	Gly		Pro	Gly	Pro	Val	Gly	Ala
30					245					250					255	
a	Ala	Gly	Ala	Thr 260	Gly	Ala	Arg	Gly	Leu 265	Val	Gly	Glu	Pro	Gly 270	Pro	Ala
35				200					203					270		
<i>.</i>	Gly	Ser	Lys 275	Gly	Glu	Ser	Gly	Asn 280	Lys	Gly	Glu	Pro	Gly 285	Ser	Ala	Gly
<b>40</b> -																
40	Pro	Gln 290	Gly	Pro	Pro	Gly	Pro 295	Ser	Gly	Glu	Glu	Gly 300	Lys	Arg	Gly	Pro
	•	<b>~1</b>	<b>61</b>	•1-	<b>~</b> 1		_ •		_	_						
45	305	GIÀ	GIU	Ala	GIÀ	310	Ala	Gly	Pro	Pro	Gly 315	Pro	Pro	Gly	Leu	Arg 320
	Gl v	Se~	Dro	G) v	Cc=	<b>7.~~</b>	<b>01</b>	7	Dw	al.	<b>5</b> 7 -	<b>.</b>	<b>63</b>	•		<b>~</b> 3
<b>50</b>	GTÅ	Ser	FLU	GIÀ	325	Arg	αтλ	Leu	LLO	330 GIA	ALA	Asp	GIÀ	Arg	Ala 335	GTA

	Val	Met	Gly	Pro	Pro	Gly	Ser	Arg	Gly	Ala	Ser	Gly	Pro	Ala	Gly	Val
5				340					345					350		
										•			_			
	Arg	Gly	Pro	Asn	Gly	Asp	Ala	_	Arg	Pro	Gly	Glu		GLY	Leu	Met
10			355					360					365			
	C1v	Pro	Arg	Glv	T.e.u	Pro	Glv	Ser	Pro	Glv	Asn	Ile	Glv	Pro	Δla	Glv
	Gry	370	,	01,			375	-		,		380	7			O <sub>L</sub>
15		3.0														
:	Lys	Glu	Gly	Pro	Val	Gly	Leu	Pro	Gly	Ile	Asp	Gly	Arg	Pro	Gly	Pro
20	385					390					395					400
20																
	Ile	Gly	Pro	Ala	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Asn	Ile	Gly	Phe	Pro
					405					410					415	
25																
•	Gly	Pro	Lys	Gly	Pro	Thr	Gly	Asp	Pro	Gly	Lys	Asn	Gly	Asp	Lys	Gly
				420					425					430		
30																
	His	Ala	Gly	Leu	Ala	Gly	Ala	Arg	Gly	Ala	Pro	Gly	Pro	Asp	Gly	Asn
			435					440					445			
35																
	Asn	•	Ala	Gln	Gly	Pro		Gly	Pro	Gln	Gly		Gln	Gly	Gly	Lys
v.		450					455					460				
40		•			_			_	_		<b></b>	<b>a</b> 1	<b>61.</b>		_	<b>~</b> 3
	•	Glu	Gln	GIÅ	Pro		GIÀ	Pro	Pro	GIÀ		GIU	GTÅ	ren	Pro	•
	465					470					475					480
45	Pro	Ser	Gly	Pro	Δla	Glv	Glu	Val	Glv	Lvs	Pro	Glv	Glu	Ara	Glv	Len
•	FIO	501	017		485	O.J	014	· · · ·	O.J	490		0-7			495	
					-03											
50	His	Glv	Glu	Phe	Glv	Leu	Pro	Glv	Pro	Ala	Gly	Pro	Ara	Gly	Glu	Arg
	,			500	- 4				505		- 4	-	,	510		•

$\vec{x}_{i}$	Gly	Pro	Pro	Gly	Glu	Ser	Gly	Ala	Ala	Gly	Pro	Thr	Gly	Pro	Ile	Gly
5			515					520					525			
•																
	Ser	Arg	Gly	Pro	Ser	Gly	Pro	Pro	Gly	Pro	Asp	Gly	Asn	Lys	Gly	Glu
10		530					535					540				
	Pro	Gly	Val	Val	Gly	Ala	Val	Gly	Thr	Ala	Gly	Pro	Ser	Gly	Pro	Ser
•	545					550					555					560
15																
,	Gly	Leu	Pro	Gly	Glu	Arg	Gly	Ala	Ala	Gly	Ile	Pro	Gly	Gly	Lys	Glv
n 1	•			-	565	_	•			570			•	•	575	1
20																
	Glu	Lvs	Glv	Glu	Pro	Glv	Leu	Ara	Glv	Glu	Ile	Glv	Asn	Pro	Gly	Ara
·		-2-	2	580		1		5	585			7		590	<b>01</b>	9
25	•								505					330		
	Δen	Glv	Δla	Δra	Glv	Δla	Wie	G) v	<b>11</b> 2	V=1	Glv	ב 1 מ	Pro	Glv	Pro	71-
	,,,op	<b>-</b> 2	595		<b>-</b>			600			<b>0.1</b>		605	Gry	110	ALG
1.			,,,					000					003			
30	C1.	775	ጥኮ~	Gl <sub>32</sub>	) en	N.c.	@1.v	<i>ر</i> ا،،	. הות	C3	מומ	712	Cl.	Dwa	Ala	<b>~</b> 3
	GIY	610	1111	Gry	мър	ALG	615	GIU	Ala	GIY	Ala		GIY	PIO	Ala	GIY
		910					013					620				
35	D===	210	C)	Dwo	N	<b>~1</b>	Com	D-10	<b>61</b>	<b>~1</b>	3	<b>~</b> 3	<b>~</b> 1	17-3	<b>~</b> 1	<b>5</b>
•		ATA	GIY	PIO	ALG		Ser	PIO	GIY	GIU		GIY	GIU	vai	Gly	
	625					630					635					640
40		~1	<b>5</b>	•	<b>~</b> 1	-1			_							_
	Ala	GIÀ	Pro	Asn		Phe	Ala	Gly	Pro		GIÅ	Ala	Ala	Gly	Gln	Pro
					645					650					655	
						_		_					_			
45	Gly	Ala	Lys		Glu	Arg	Gly	Ala		Gly	Pro	Lys	Gly	Glu	Asn	Gly
				660					665					670		
50	Val	Val		Pro	Thr	Gly	Pro	Val	Gly	Ala	Ala	Gly	Pro	Ala	Gly	Pro
			675					680					685			

5	Asn	_	Pro	Pro	Gly	Pro		Gly	Ser	Arg	Gly	Asp	Gly	Gly	Pro	Pro
÷		690					695					700				
10	•	Met	Thr	Gly	Phe		Gly	Ala	Ala	Gly	_	Thr	Gly	Pro	Pro	•
5	705					710					715					720
15	Pro	Ser	Gly	Ile	Ser 725	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly 735	Lys
					123					730					733	
20	Glu	Gly	Leu	Arg 740	Gly	Pro	Arg	Gly	Asp 745	Gln	Gly	Pro	Val	Gly 750	Arg	Thr
20																
<u>.</u> .	Gly	Glu	Val 755	Gly	Ala	Val	Gly	Pro 760	Pro	Gly	Phe	Ala	Gly 765	Glu	Lys	Gly
25																
,	Pro	<b>Ser</b> 770	Gly	Glu	Ala	Gly	775	Ala	Gly	Pro	Pro	Gly 780	Thr	Pro	Gly	Pro
30																
:	Gln 785	Gly	Leu	Leu	Gly	790	Pro	Gly	Ile	Leu	Gly 795	Leu	Pro	Gly	Ser	Arg
35																
**	Gly	Glu	Arg	Gly	Leu 805	Pro	Gly	Val	Ala	Gly 810	Ala	Val	Gly	Glu	Pro 815	Gly
40	Dwa	Lou	G) v	Tla	71.	G) v	Dro	Dra	G) v	λla	7 × 4	Gly	Pro	Pro	elv.	<b>N1</b> a
	PIO	Deu	Gly	820	714	Gly	210	PIO	825	A14	ALG	GLY	710	830	GLY	AL a
45	Val	Glv	Ser	Pro	Gly	Val	Asn	Glv	Ala	Pro	Gly	Glu	Ala	Gly	Arq	qeA
			835					840					845	4	- 3	-£
50	Gly	Asn	Pro	Gly	Asn	Asp	Gly	Pro	Pro	Gly	Arg	Asp	Gly	Gln	Pro	Gly
		850					855					860				

55

3:

<i>4</i>	His	Lys	Gly	Glu	Arg	Gly	Tyr	Pro	Gly	Asn	Ile	Gly	Pro	Val	Gly	Ala
5	865					870					875					880
÷	Ala	Gly	Ala	Pro	Gly	Pro	His	Gly	Pro	Val	Gly	Pro	Ala	Gly	Lys	His
10					885					890					895	
*1	Gly	Asn	Arg	Gly	Glu	Thr	Gly	Pro	Ser	Gly	Pro	Val	Gly	Pro	Ala	Gly
15				900					905					910		
,	Ala	Val	Gly	Pro	Arg	Gly	Pro	Ser	Gly	Pro	Gln	Gly	Ile	Arg	Gly	Asp
20			915					920					925			
	Lys	Gly	Glu	Pro	Gly	Glu	Lys	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Leu	Lys
25		930					935					940				
	Gly	His	Asn	Gly	Leu	Gln	Gly	Leu	Pro	Gly	Ile	Ala	Gly	His	His	Gly
•	945					950					955					960
30	A en	Gl n	Glv	Ala	Pro	Glv	Ser	17a 1	Gly	Pro	בומ	GI v	Pro	7~~	<b>71.</b>	Dwa
÷	vob	GI.II	Gry	nza	965	GLY	361	val	GIY	970	nia	Gly	FIO	AT.	975	PIO
35	Ala	Gly	Pro	Ser	Gly	Pro	Ala	Gly	Lys	Asp	Gly	Arg	Thr	Gly	His	Pro
5				980	-			-	985	-	•			990		
40	Gly	Thr	Val	Gly	Pro	Ala	Gly	Ile	Arg	Gly	Pro	Gln	Gly	His	Gln	Gly
			995					1000					1005	5		
45	Pro	Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Val
		1010	)				1015	5	٠			1020	)			٠
50			Gly	Gly	Tyr			Gly	Tyr	qaA	Gly	Asp	Phe	Tyr	Arg	Ala
	102	5				1030	)				1039	5				1040

#### (2) INFORMATION FOR SEQ ID NO:31:

5

#### (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3120 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

10

#### (ii) MOLECULE TYPE: cDNA

20

25

30

35

40

45

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CAGTACGACG GTAAAGGCGT AGGCCTGGGT CCGGGTCCGA TGGGCCTGAT GGGTCCACGT 60 GGCCCACCGG GTGCAGCAGG TGCGCCGGGT CCGCAGGGCT TCCAAGGTCC GGCGGGTGAA CCGGGCGAAC CGGGTCAGAC GGGTCCGGCG GGTGCTCGCG GTCCGGCTGG CCCACCGGGC 180 AAAGCTGGCG AAGACGGTCA CCCGGGTAAG CCAGGCCGCC CGGGCGAACG TGGCGTCGTG 240 GGTCCGCAAG GTGCGCGTGG TTTCCCGGGC ACGCCGGGTC TGCCGGGTTT CAAAGGCATT 300 CGTGGTCACA ACGGTCTGGA CGGTCTGAAA GGCCAACCGG GTGCTCCGGG CGTCAAAGGC 360 GAACCGGGTG CCCCAGGCGA AAACGGTACG CCGGGCCAGA CTGGTGCGCG TGGTCTGCCG 420 GGTGAACGCG GCCGTGTTGG CGCTCCGGGT CCGGCTGGCG CGCGTGGCAG CGATGGCTCC 480 GTCGGTCCGG TTGGCCCTGC GGGTCCGATT GGTTCCGCTG GCCCTCCGGG TTTCCCGGGT 540 GCGCCGGGTC CGAAGGGTGA GATCGGCGCG GTTGGCAACG CAGGCCCGGC TGGTCCAGCC 600 GGCCCTCGTG GCGAAGTCGG TCTGCCGGGT CTGAGCGGTC CGGTAGGCCC ACCGGGTAAC 660

\$ 0

720	CGTTGCCGGT	GCCTGCCGGG	GGTGCGGCTG	GGGTGCAAAA	ACGGCCTGAC	CCGGGCGCAA	
780	CGGTGCAACT	TAGGCGCAGC	CCGGGTCCGG	GCGCGGTATT	TGCCGGGTCC	GCCCGGGCC	
840	AAGCGGTAAC	CTAAAGGCGA	CCGGCGGGTT	CGAACCGGGT	GCCTGGTTGG	GGTGCCCGTG	
900	CGAAGAAGGT	GTCCGAGCGG	GGTCCGCCGG	GGGCCCGCAG	CGGGTTCCGC	AAAGGTGAGC	
960	GGGTCTGCGT	CGGGTCCGCC	GCAGGCCCTC	GGCTGGTTCC	CGAACGGCGA	AAACGTGGTC	
1020	GATGGGTCCG	GTGCGGGCGT	GCGGACGGCC	CCTGCCGGGC	GTAGCCGTGG	GGCAGCCCGG	
1080	CGACGCGGGC	GTCCGAATGG	GGTGTCCGTG	TGGTCCGGCT	GTGGTGCCTC	CCGGGTTCCC	
1140	GGGTAACATT	CGGGTAGCCC	CGTGGCCTGC	GATGGGTCCG	AACCGGGCCT	CGTCCGGGTG	
1200	TCCGGGTCCG	TTGATGGTCG	CTGCCGGGTA	TCCGGTAGGT	GTAAGGAGGG	GGTCCGGCGG	
1260	TCCGAAGGGT	GTTTTCCGGG	GGTAACATCG	TGGCGAGCCG	CGGGCGCTCG	ATCGGCCCTG	
1320	AGGTGCCCGT	CAGGTCTGGC	AAAGGCCATG	GAACGGTGAT	ACCCGGGCAA	CCGACGGGCG	
1380	GCAGGGCGTA	CGCCGGGTCC	GCGCAGGGTC	TAACAATGGT	GTCCGGATGG	GGTGCACCGG	
1440	TCTGCCGGGT	GCTTCCAGGG	GGCCCACCGG	GGGTCCGGCA	AAGGTGAACA	CAGGGTGGCA	
1500	TGGCGAGTTT	GTGGCCTCCA	CCGGGCGAAC	AGTGGGCAAA	CGGCTGGTGA	CCGAGCGGCC	
1560	ATCCGGCGCG	CTCCGGGCGA	GAGCGCGGCC	TCCGCGTGGT	GTCCGGCCGG	GGCCTGCCGG	
1620	TCCGGACGGC	GCCCACCGGG	GGTCCGAGCG	TGGTTCCCGT	CCGGCCCGAT	GCAGGTCCGA	
1680	TGGTCCGAGC	CCGGCCCGTC	GTTGGTACCG	TGTTGGTGCT	AGCCGGGTGT	AACAAAGGCG	

· 5

GGTCTGCCGG	GCGAACGCGG	TGCCGCTGGT	ATTCCGGGCG	GCAAAGGTGA	AAAAGGTGAA	1740
CCGGGTCTGC	GCGGTGAGAT	TGGCAACCCG	GGCCGTGAÇG	GTGCTCGCGG	TGCACACGGC	1800
GCGGTTGGCG	CACCGGGTCC	GGCAGGCGCG	ACTGGTGATC	GTGGCGAAGC	TGGTGCAGCG	1860
GGTCCGGCGG	GTCCGGCCGG	CCCTCGCGGT	TCCCCGGGCG	AACGCGGCGA	AGTCGGCCCG	1920
GCTGGCCCGA	ATGGCTTTGC	TGGCCCAGCG	GGCGCTGCGG	GCCAACCGGG	TGCGAAAGGT	1980
GAGCGCGGTG	CCAAAGGCCC	GAAAGGTGAA	AATGGTGTAG	TTGGTCCGAC	GGGTCCGGTT	2040
GGTGCGGCTG	GTCCGGCTGG	CCCGAATGGT	CCGCCGGGTC	CGGCAGGCAG	CCGTGGCGAT	2:100
GGTGGCCCAC	CGGGCATGAC	CGGTTTCCCT	GGCGCGGCCG	GTCGCACCGG	CCCGCCGGGT	2160
CCGTCTGGCA	TTTCTGGCCC	ACCGGGTCCG	CCGGGTCCGG	CGGGCAAAGA	AGGTCTGCGT	2220
GGCCCACGCG	GCGACCAGGG	TCCGGTGGGC	CGTACCGGCG	AAGTCGGTGC	TGTTGGCCCT	2280
CCGGGCTTTG	CGGGTGAGAA	AGGTCCGAGC	GGTGAAGCTG	GCACCGCAGG	CCCGCCGGGT	2340
ACGCCGGGTC	CGCAAGGTCT	GCTGGGTGCT	CCGGGTATCC	TGGGCCTGCC	GGGCTCCCGT	2400
GGCGAACGCG	GTCTGCCGGG	CGTTGCAGGC	GCTGTAGGCG	AACCGGGTCC	GCTGGGTATC	2460
GCGGGTCCGC	CGGGTGCGCG	TGGTCCGCCG	GGTGCCGTGG	GCTCTCCGGG	TGTTAACGGC	2520
GCCCCTGGTG	AAGCGGGCCG	CGACGGCAAT	CCGGGCAACG	ATGGTCCGCC	GGGTCGTGAT	2580
GGTCAGCCGG	GTCACAAAGG	TGAGCGTGGC	TACCCGGGTA	ACATCGGTCC	GGTTGGTGCG	2640
GCCGGCGCTC	CGGGTCCGCA	CGGTCCGGTA	GGCCCAGCCG	GCAAACACGG	TAACCGTGGT	2700

.: 5

£.

.:	GAAACGGGTC CGTCCGGTCC GGTAGGTCCG GCGGGTGCTG TTGGTCCACG CGGCCCGTCC 2	760
5	GGCCCGCAGG GTATTCGCGG TGACAAAGGC GAACCGGGCG AAAAAAGGTCC GCGTGGTCTG 20	820
10	CCGGGCCTTA AGGGCCACAA CGGTCTGCAA GGTCTGCCGG GTATCGCGGG TCACCACGGT 20	880
d.	GATCAGGGTG CTCCGGGTTC CGTTGGTCCG GCCGGTCCGC GTGGCCCGGC TGGTCCGTCT 29	940
15	GGTCCGGCCG GTAAAGACGG CCGTACGGGC CACCCGGGTA CGGTGGGTCC GGCCGGCATT 3	000
	CGCGGTCCGC AAGGTCACCA GGGTCCGGCG GGTCCGCCGGG TCCGCCGGGT 3	060
	CCGCCGGGTG TTAGCGGTGG CGGTTATGAT TTTGGTTATG ACGGTGATTT CTATCGTGCG 3:	120
25		
	(2) INFORMATION FOR SEQ ID NO:32:	
;	(i) SEQUENCE CHARACTERISTICS:	
30	(A) LENGTH: 1040 amino acids	
•	(B) TYPE: amino acid	
	(C) STRANDEDNESS: single	
35	(D) TOPOLOGY: unknown	
:	(ii) MOLECULE TYPE: peptide	
40		
•		
45	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:	
50·	Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu  1 5 10 15	
4		

ı:	Met	Gly	Pro	Arg	Gly	Pro	Pro	Gly	Ala	Ala	Gly	Ala	Pro	Gly	Pro	Gln
5				20					25					30		
i. 10	Gly	Phe	Gln 35	Gly	Pro	Ala	Gly	Glu 40	Pro	Gly	Glu	Pro	Gly 45	Gln	Thr	Gly
	Pro	Ala	Gly	Ala	Arg	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Lys	Ala	Gly	Glu
15		50					55					60				
	Asp 65	Gly	His	Pro	Gly	Lys 70	Pro	Gly	Arg	Pro	Gly 75	Glu	Arg	Gly	Val	Val 80
20	Gly	Pro	Gln	Gly	Ala	Arg	Gly	Phe	Pro	Gly	Thr	Pro	Gly	Leu	Pro	Gly
25					85			_		90			_	_	95	
.a	Phe	Lys	GIÀ	11e 100	Arg	GIĀ	H1s	Asn	105	Leu	Asp	GTÅ	Leu	110	GIĄ	GIn
30	Pro	Gly	Ala 115	Pro	Gly	Val	Lys	Gly 120	Glu	Pro	Gly	Ala	Pro	Gly	Glu	Asn
35 <sup>,</sup>	Gly	Thr		Gly	Gln	Thr	Gly		Arg	Gly	Leu	Pro	Gly	Glu	Arg	Gly
•		130					135					140				
40	Arg 145	Val	Gly	Ala	Pro	<b>Gly</b> 150	Pro	Ala	Gly	Ala	Arg 155	Gly	Ser	Asp	Gly	Ser 160
45	Val	Gly	Pro		Gly 165		Ala	Gly	Pro	Ile 170	Gly	Ser	Ala	Gly	Pro 175	Pro
50	Gly	Phe	Pro		Ala	Pro	Gly	Pro	_	Gly	Glu	Ile	Gly		Val	Gly
				180					185					190		

*[*.

_	Asn	Ala	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Glu	Val	Gly	Leu
5			195					200					205			
	Pro	Gly	Leu	Ser	Gly	Pro	Val	Gly	Pro	Pro	Gly	Asn	Pro	Gly	Ala	Asn
10		210					215					220				
	Gly	Leu	Thr	Gly	Ala		Gly	Ala	Ala	Gly		Pro	Gly	Val	Ala	Gly
15	225					230					235					240
		_		_			_	_		-1			_			
	Ala	Pro	Gly	Leu	Pro	Gly	Pro	Arg	Gly		Pro	GIÀ	Pro	Val	•	Ala
20					245					250					255	
20	11.	<b>~1.</b> *	212	mb *	Gly	21-	N ~~	C1.,	Lou	Wa I	C1v	al	Dro	Glv.	Dwa	210
	Ala	GIY	MIG	260	GIA	MIG	Arg	GIY	265	Val	Gly	GIU	FIO	270	PIO	ATA
				200					203					2,0		
25	Glv	Ser	LVR	Glv	Glu	Ser	Glv	Asn	Lvs	Glv	Glu	Pro	Glv	Ser	Ala	Glv
	017	-	275	,			1	280	-4-	2			285			1
30	Pro	Gln	Gly	Pro	Pro	Gly	Pro	Ser	Gly	Glu	Glu	Gly	Lys	Arg	Gly	Pro
		290					295					300				
35	Asn	Gly	Glu	Ala	Gly	Ser	Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Leu	Arg
	305					310					315					320
	Gly	Ser	Pro	Gly	Ser	Arg	Gly	Leu	Pro	Gly	Ala	Asp	Gly	Arg	Ala	Gly
40					325					330					335	
	Val	Met	Gly	Pro	Pro	Gly	Ser	Arg	Gly	Ala	Ser	Gly	Pro	Ala	Gly	Val
45				340					345					350		
	Arg	Gly		Asn	Gly	Asp	Ala	-	Arg	Pro	Gly	Glu		Gly	Leu	Met
50			355					360					365			

	Gly		Arg	Gly	Leu	Pro		Ser	Pro	Gly	Asn		Gly	Pro	Ala	Gly
5		370					375					380				
	Lys	Glu	Gly	Pro	Val	Gly	Leu	Pro	Gly	Ile	Asp	Gly	Arg	Pro	Gly	Pro
10	385					390					395					400
	Ile	Gly	Pro	Ala	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Asn	Ile	Gly	Phe	Pro
15					405					410					415	
	Gly	Pro	Lys	Gly	Pro	Thr	Gly	Asp	Pro	Gly	Lys	Asn	Gly	Asp	Lys	Gly
				420					425					430		
20	His	Ala	Gly	Leu	Ala	Gly	Ala	Arg	Gly	Ala	Pro	Gly	Pro	qaA	Gly	Asn
			435			-		440	-			•	445	-	•	
25	Asn	Glv	Ala	Gln	Gly	Pro	Pro	Glv	Pro	Gln	Glv	Val	Gln	Glv	Glv	Twa
		450			•		455				,	460		1		-,,
30	Gly	G) u	Gln.	G) v	Pro	חות	Glv.	Dwa	Dwa	Gly	Dho	C1 m	<b>C</b> 1	T 0	B	<b>a</b> 1
	465	GIU	GIII	GLY	PIO	470	GIŞ	PIO	PIO	GIY	475	GIII	GIŞ	Leu	PIO	480
35	_			_				_								
	Pro	Ser	GΙΆ	Pro	A1a 485	Gly	Glu	Val	Gly	Lys 490	Pro	Gly	Glu	Arg	Gly 495	Leu
. *																
40	His	Gly	Glu	Phe 500	Gly	Leu	Pro	Gly	Pro 505	Ala	Gly	Pro	Arg	Gly 510	Glu	Arg
									303					310		
45	Gly	Pro		Gly	Glu	Ser	Gly		Ala	Gly	Pro	Thr		Pro	Ile	Gly
			515					520					525			
50	Ser		Gly	Pro	Ser	Gly		Pro	Gly	Pro	Asp		Asn	Lys	Gly	Glu
		530					535					540				

55·

	Pro	Gly	Val	Val	Gly	Ala	Val	Gly	Thr	Ala	Gly	Pro	Ser	Gly	Pro	Ser
5	545					550					555					560
:	Gly	T.A.	Pro	Glý	Glu	Ara	Gly	בות	בות	Gly	Tla	Bro	Gly	Gl v	T	Gly
	GIY	Deu	-10	Gry		ALY	GIY	AIG	AIG		116	FIO	GIY	GIY		GIY
10					565					570					575	
	Glu	Lys	Gly	Glu	Pro	Gly	Leu	Arg	Gly	Glu	Ile	Gly	Asn	Pro	Gly	Arg
15				580					585					590		
75																
	7	~1·	7.1.a	N ===	<b>~1</b>	N1 a	T74 -	<b>~1</b>	81.	17-1	~1	21.0	D	<b>G1</b>	<b>D</b>	
	Азр	GIÀ		ALG.	GIŞ	ATA	uis		AIA	val	GIY	Ala		Gly	Pro	ALA
20			595					600					605			
	Gly	Ala	Thr	Gly	Asp	Arg	Gly	Glu	Ala	Gly	Ala	Ala	Gly	Pro	Ala	Gly
		610					615					620				
25																
	Pro	פות	Gly	Pro	2 ~~	Gly	Car	Bro	C1.	Gl.	n-a	G1.v	C1	Val	<b>~1</b>	Desc
		AIG	GIY	FIO	Arg		Ser	PIO	GIY	GIU		GIY	GIU	AGT	GIA	
	625					630					635					640
30																
	Ala	Gly	Pro	Asn	Gly	Phe	Ala	Gly	Pro	Ala	Gly	Ala	Ala	Gly	Gln	Pro
;					645					650					655	
•																
35'	Glv	בומ	Tara	GIV	Glu	Ara	Gly	- 1 מ	Tara	GI v	Dro	Tara	Glv.	al	Non-	Gly
	dry	7.14	275		O.L.	<i></i> 9	GLY	AIG		Gry		цуь	GLY		ABII	GIÅ
				660					665					670		
40																
40	Val	Val	Gly	Pro	Thr	Gly	Pro	Val	Gly	Ala	Ala	Gly	Pro	Ala	Gly	Pro
			675					680					685			
45	Asn	Glv	Pro	Pro	Glv	Pro	Ala	Glv	Ser	Ara	Glv	Asn	Glv	Gly	Pro	Pro
	•	690			2			,		5	,		<b>U</b> 2.3	O <sub>1</sub>		110
		030					695					700				
50	Gly	Met	Thr	Gly	Phe	Pro	Gly	Ala	Ala	Gly	Arg	Thr	Gly	Pro	Pro	Gly
	705					710					715					720

	Pro	Ser	Gly	Ile	Ser	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Lys
5					725					730					735	
										•						
;	Glu	Gly	Leu	Arg	Gly	Pro	Arg	Gly	Asp	Gln	Gly	Pro	Val	Gly	Arg	Thr
10				740					745					750	·	
÷	Gly	Glu	Val	Gly	Ala	Val	Gly	Pro	Pro	Gly	Phe	Ala	Gly	Glu	Lys	Gly
15			755					760					765			
	Pro	Ser	Gly	Glu	Ala	Gly	Thr	Ala	Gly	Pro	Pro	Gly	Thr	Pro	Gly	Pro
20		770					775					780				
		Gly	Leu	Leu	Gly	Ala	Pro	Gly	Ile	Leu	Gly	Leu	Pro	Gly	Ser	Arg
25	785					790					795					800
20																
	Gly	Glu	Arg	Gly		Pro	Gly	Val	Ala		Ala	Val	Gly	Glu	Pro	Gly
					805					810					815	
30	_	_	-1	-1.			_	_					_	_		
	Pro	Leu	GIÀ		Ата	GIĀ	Pro	Pro	_	Ala	Arg	GIY	Pro		Gly	Ala
:				820					825					830		
35	77.0.7	<b>~1</b>	50 <b>m</b>	Dwa	<b>a</b> 1	37a 1	<b>3</b>	<b>01</b>	23.	Dua	a1	<b>~1</b>	21-	<b>03</b>	•	
	vai	GIY	835	PIO	GIŞ	Val	ASII	840	ALA	PIO	GIŞ	GIU	845	GIÀ	Arg	Asp
9			033					040					043			
40	Glv	Asn	Pro	Glv	Asn	Asp	Glv	Pro	Pro	G] v	Ara	Asn	Glv	Gln	Pro	Gly
		850		2			855			,	3	860	,			O.
45	His	Lys	Gly	Glu	Arg	Gly	Tyr	Pro	Gly	Asn	Ile	Gly	Pro	Val	Gly	Ala
	865	-			•	870	-		•		875	•			•	880
50	Ala	Gly	Ala	Pro	Gly	Pro	His	Gly	Pro	Val	Gly	Pro	Ala	Gly	Lys	His
					885			-		890	=			-	895	
									•							

	Gly	Asn	Arg	Gly	Glu	Thr	Gly	Pro	Ser	Gly	Pro	Val	Gly	Pro	Ala	Gly
5				900					905					910		
	Ala	Val	Gly	Pro	Arg	Gly	Pro	Ser	Gly	Pro	Gln	Gly	Ile	Arg	Gly	Asp
10			915					920					925			_
	Lvs	Gly	Glu	Pro	Glv	Glu	Lvs	Glv	Pro	Ara	Glv	Leu	Pro	Glv	T.011	Tara
	7	930			,		935	,			,	940		ory	ыeu	пåа
15																
		His	Asn	Gly	Leu		Gly	Leu	Pro	Gly		Ala	Gly	His	His	Gly
20	945					950					955					960
	Asp	Gln	Gly	Ala	Pro	Gly	Ser	Val	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Pro
					965					970					975	
25			_	_				_			_					
	Ala	Gly	Pro	980	Gly	Pro	Ala	Gly	1985 1985	Asp	Gly	Arg	Thr	Gly 990	His	Pro
				500					<i>5</i> 05					330		
30	Gly	Thr	Val	Gly	Pro	Ala	Gly	Ile	Arg	Gly	Pro	Gln	Gly	His	Gln	Gly
			995					1000	)				1005	5		
35	Pro	Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Glv	Val
		1010				-	1015		•			1020			2	
40	Ser 1025	Gly	Gly	Gly	Tyr	Asp 1030		Gly	Tyr	Asp	Gly 1035		Phe	Tyr	Arg	
:	1023					1030	,				1035	,				1040
45																
	(2) INFOR	LTAMS	ON F	FOR S	EQ I	D NC	:33:									
	(i)	SEQU	JENCF	сна	RACT	ERTS	TICS	l :								
50	,_,		LEN													
		(B)	TYP	'Е: п	ucle	eic a	cid									
		(C)	STR	ANDE	DNES	SS: s	ingl	.e								
55																

5	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:	
ς.	GGAATTCATG CAGTATGATG GCAAAGGCGT CGGCCTCGGC CCGGGCCCAA TGGGCCTCAT	60
	GGGCCCGCGC GGCCCA	76
20	(2) INFORMATION FOR SEQ ID NO:34:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 79 base pairs	
25	(B) TYPE: nucleic acid	
•	(C) STRANDEDNESS: single	
*	(D) TOPOLOGY: linear	
30		
*	(ii) MOLECULE TYPE: cDNA	
ti.		
35		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:	
40	CCGGGCGCGC CGGGTGGCCC ACGTCGACCG CGGGGTCCGG GCGTTCCAAA GGTCCCGGGA	60
45	CGGCCAATTA TTCGAACCC	79
	(2) INFORMATION FOR SEQ ID NO:35:	
50	(i) SEQUENCE CHARACTERISTICS:	•
	(A) LENGTH: 82 base pairs	
•	(B) TYPE: nucleic acid	
55		

5 ·	(c) SIRMIDEDITESS: SINGIE	
5.	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
10		
·	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:	
·15·	GGAATTCGCC GGTGAGCCGG GTGAACCGGG CCAAACGGGT CCGGCAGGTC CACGTGGTCC	60
	AGCGGGCCCG CCTGGCAAGG CG	82
20	(2) INFORMATION FOR SEQ ID NO:36:	
	(i) SEQUENCE CHARACTERISTICS:	
25	(A) LENGTH: 84 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
30		
	(ii) MOLECULE TYPE: cDNA	
35·		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:	
40 <sup>-</sup>	(012) 012	
	CCGGGCGGAC CGTTCCGCCC ACTTCTACCG GTGGGACCGT TTGGCCCGGC GGGCCACTCG	60
	coodean corrected nertenece crowneed rightened addedcrees	00
45 <sup>.</sup>	CACCGCATCA CATTATTCGA ACCC	84
	CACCOCATCA CATATIONA ACCC	04
	(2) INFORMATION FOR SEQ ID NO:37:	
50		
-•	(i) SEQUENCE CHARACTERISTICS:	
÷	(A) LENGTH: 240 base pairs	
55		

(B) TYPE: nucleic acid

3	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
10	(ii) MOLECULE TYPE: cDNA	
:	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:	
15	CAGTATGATG GCAAAGGCGT CGGCCTCGGC CCGGGCCCAA TGGGCCTCAT GGGCCCGCGC	60
20	GGCCCACCGG GTGCAGCTGG CGCCCCAGGC CCGCAAGGTT TCCAGGGCCC TGCCGGTGAG	120
	CCGGGTGAAC CGGGCCAAAC GGGTCCGGCA GGTGCACGTG GTCCAGCGGG CCCGCCTGGC	180
25	AAGGCGGGTG AAGATGGCCA CCCTGGCAAA CCGGGCCGCC CGGGTGAGCG TGGCGTAGTG	240
30	(2) INFORMATION FOR SEQ ID NO:38:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 80 amino acids	
35	(B) TYPE: amino acid	-
	(C) STRANDEDNESS: single (D) TOPOLOGY: unknown	
40		
	(ii) MOLECULE TYPE: peptide	
45		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:	
50	Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu	
	1 5 10 15	
55		

5	Met	Gly	Pro		Gly	Pro	Pro	Gly		Ala	Gly	Ala	Pro		Pro	Gln	
·;				20					25					30			
10	Gly	Phe	Gln 35	Gly	Pro	Ala	Gly	Glu 40	Pro	Gly	Glu	Pro	Gly 45	Gln	Thr	Gly	
ı																	
15	Pro	Ala 50	Gly	Ala	Arg	Gly	Pro 55	Ala	Gly	Pro	Pro	Gly 60	Lys	Ala	Gly	Glu	
75		30					23					00					
	Asp	Gly	His	Pro	Gly	Lys	Pro	Gly	Arg	Pro	Gly	Glu	Arg	Gly	Val	Val	
20	65					70					75					80	
<b>,</b>																	
25	(2) INFO	RMAT:	ION :	FOR S	SEQ :	ID N	0:39	:									
••	(i)	SEQ	UENC	E CHI	ARAC'	reri:	STIC	S:									
		(A)	) LEI	NGTH	: 27	baa	ве ра	airs									
30		(B)	TY:	PE: 1	nucl	eic a	acid										
r <u>'</u>		(C)	) ST	RANDI	EDNE	SS: s	sing:	le									
		(D)	) TO	POLO	BY:	linea	ar										
35	, (ii)	MOL	ECULI	E TY	PE: (	cDNA											
40	(xi)	SEQ	UENC	E DES	SCRI	PTIO	N: SI	EQ II	ONO:	:39:							
·	ATGGGGCT	CG C	rggc	CCAC	C GG(	GCGAI	ACCG	GGT	CCGC	CAG (	3CCC(	AAAE	G T	CGC	GTGG(	2	60
45	GATAGCGG	GC T	CGCT	GCC(	C AC	CGGG	CGAA	CCG	GTC	CGC (	CAGG	CCG	AA AG	GTC	CGCG	r	120
50	GGCGATAG	CG G(	GCTC	3CTG(	3 CC	CACC	3GGC	GAA	CGGG	GTC (	CGCC/	AGGC	CC G	<b>AAA</b> G(	GTCC	3	180
%	CGTGGCGA	TA G	CGGG	CTCG	C TG	GCC()	ACCG	GGC	BAAC	CGG (	GTCC(	GCCA	G C	CCGAI	AAGG7	<b>F</b>	240
55	CCGCGTGG	CG A	TAGC	3GGC'	r cc	CGGG	CGAT	TCC	AA1								276

	(4)	INFOR	UMAI	LOM I	OR	BQ.	LD IN	J. 40	•								
5																	
.,,		(i)	SEQU	JENCI	E CH	ARAC:	reri:	STICS	3:								
••			(A)	LE	NGTH	: 91	amin	no a	cids								
10			(B)	TYI	?E: 8	amino	ac:	id									
			(C)	STI	RANDI	EDNE	SS: 8	sing	le								
er -			(D)	TOI	OTO	3Y: 1	ınkno	own									
n																	
15		(ii)	MOL	CULI	TYI	PE: 1	pept:	ide									
										•							
		(xi)	SEQU	JENCI	E DES	CRI	PTIO	<b>1:</b> SI	EQ II	ои с	:40:						
20																	
		Met	Gly	Leu	Ala	Glý	Pro	Pro	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Lys
		1				5					10					15	
25																	
		Gly	Pro	Arg	Gly	Asp	Ser	Gly	Leu	Ala	Gly	Pro	Pro	Gly	Glu	Pro	Gly
					20					25					30		
30																	
		Pro	Pro	Gly	Pro	Lys	Gly	Pro	Arg	Gly	Asp	Ser	Gly	Leu	Ala	Gly	Pro
*				35					40					45			
•																	
35		Pro	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Lys	Gly	Pro	Arg	Gly	qaA	Ser
j			50					55					60				
•					_												
40			Leu	Ala	Gly	Pro		Gly	Glu	Pro	Gly		Pro	Gly	Pro	Lys	Gly
		65					70					75					80
			_	<b>~</b> 3.		_			_		_	_					
45		Pro	Arg	GIÀ	qaa		GLY	Leu	Pro	GIÀ	_	Ser					
						85					90						

(2) INFORMATION FOR SEQ ID NO:41:

	(2)
5	
	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 13 amino acids
	(B) TYPE: amino acid
10	(C) STRANDEDNESS: single
	(D) TOPOLOGY: unknown
<i>:1</i>	
15	(ii) MOLECULE TYPE: peptide
.*	(22)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:
20	(A1) D18D101 1100000 012 00 00001
20	Gly Pro Pro Gly Leu Ala Gly Pro Pro Gly Glu Ser Gly
	1 5 10
25	(2) INFORMATION FOR SEQ ID NO:42:
	(2) INFORMATION FOR SEQ 12 NO. 42.
	(i) SEQUENCE CHARACTERISTICS:
30	(A) LENGTH: 13 amino acids
•	(B) TYPE: amino acid
	(C) STRANDEDNESS: single
35	(D) TOPOLOGY: unknown
	(b) 10201001. unuiowii
	(ii) MOLECULE TYPE: peptide
Ų.	(11) Normedan III n. populac
40	
	(ix) FEATURE:
	(A) NAME/KEY: Modified-site
45	(B) LOCATION: 23
	(D) OTHER INFORMATION: /product= "4-hydroxyproline"
	(b) other information: /products 4-nydroxyprofine
50	

	(IX) FEATURE:	
. <del>:</del> 5	(A) NAME/KEY: Modified-site	
J	(B) LOCATION: 89	
	(D) OTHER INFORMATION: /product= "Xaa = 4-hydroxyproline"	
3		
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:	
	Gly Xaa Xaa Gly Leu Ala Gly Xaa Xaa Gly Glu Ser Gly	
15	1 5 10	
20	(2) INFORMATION FOR SEQ ID NO:43:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 660 base pairs	
	(B) TYPE: nucleic acid	
25	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
,	(b) Ioloboti. Ilmeal	
30	(ii) MOLECULE TYPE: CDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:	
35	ATGGGCCCGC CGGGTCTGGC GGGCCCTCCG GGTGAAAGCG GTCGTGAAGG CGCGCCGGGT	60
40	GCCGAAGGCA GCCCAGGCCG CGACGGTAGC CCGGGGGCCA AAGGGGATCG TGGTGAAACC	120
	GGCCCGGCGG GCCCCCCGGG TGCACCGGGC GCGCCGGGTG CCCCAGGCCC GGTGGGCCCG	180
45	GCGGGCAAAA GCGGTGATCG TGGTGAGACC GGTCCGGCGG GCCCGGCCGG TCCGGTGGGC	240
50	CCAGCGGGCG CCCGTGGCCC GGCCGGTCCG CAGGGCCCGC GGGGTGACAA AGGTGAAACG	300
	GGCGAACAGG GCGACCGTGG CATTAAAGGC CACCGTGGCT TCAGCGGCCT GCAGGGTCCA	360

	CCGGGCCC	GC C	GGGC	AGTC	: GG(	GTGA	ACAG	GGT	CCGT	CCG	GAGC	CAGC	GG G	CCGG	CGGG	<b>c</b> .	420
5	CCACGCGG	TC C	GCCG(	GCAG	G CG	CGGG	CGCG	CCG	GCA	<b>A</b> ĄG	ACGG'	rctg:	AA C	GGTC	TGCC	3	480
10.	GGCCCGAT	CG G	CCCG	CCGGG	CC	CACG	CGGC	CGC	ACCG	GTG	ATGC	GGT(	CC G	GTGG	GTCC	2	540
3	CCGGGCCC	GC C	GGGC	cccc	: AG	GCCC	GCCG	GGA	CCGC	CGA	GCGC	GGGT"	rt c	GACT'	TCAG	2	600
15	TTCCTGCC	GC A	GCCG	CCGCF	. GG	AGAA	AGCG	CAC	GACGO	GCG	GTCG	CTAC'	ra c	CGTG	CGTA	A	660
20	(2) INFO	RMAT:	ION 1	FOR S	EQ :	ID N	0:44	:									
	(i)	SEQ	UENC	E CHA	RAC'	TERI:	STICS	S :									
25		(A)	) LEI	NGTH :	21	9 am:	ino a	acid	3								
	(B) TYPE: amino acid (C) STRANDEDNESS: single																
	(C) STRANDEDNESS: single (D) TOPOLOGY: unknown																
<b>30</b> .	(D) TOPOLOGY: unknown																
9	(ii)	MOL	ECULI	E TYP	E: ]	pept	ide										•
35	(xi)	SEQ	UENCI	E DES	CRI	PTIO	N: SI	EQ II	OM C	:44:							
	Met	Gly	Pro	Pro	Gly	Leu	Ala	Gly	Pro	Pro	Gly	Glu	Ser	Gly	Arq	Glu	
40	1				5			_		10	_			-	15		
•	Gly	Ala	Pro	Gly	Ala	Glu	Gly	Ser	Pro	Gly	Arg	Asp	Gly	Ser	Pro	Gly	
45				20					25					30			
	Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Ala	
50			35		-	_		40	-			-	45		•		
50																	
	Pro		Ala	Pro	Gly	Ala		Gly	Pro	Val	Gly		Ala	Gly	Lys	Ser	
55		50					55					60					

		Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro	Val	Gly
5	*	65					70				•	75					80
		Pro	Ala	Gly	Ala	Arg	Gly	Pro	Ala	Gly	Pro	Gln	Gly	Pro	Arg	Gly	Asp
10						85					90					95	
		Lys	Gly	Glu	Thr	Gly	Glu	Gln	Gly	Asp	Arg	Gly	Ile	Lys	Gly	His	Arg
15					100					105					110		
		Gly	Phe	Ser	Gly	Leu	Gln	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Ser	Pro	Gly
20				115					120					125			
		Glu	Gln	Gly	Pro	Ser	Gly	Ala	Ser	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Pro
			130					135					140				
25		Pro	Glv	Ser	Ala	Glv	Ala	Pro	Glv	Lvs	Asp	Glv	Leu	Asn	Glv	Leu	Pro
		145	O <sub>1</sub>	501			150		,	-,-		155		******			160
 <b>30</b> .																	
		Gly	Pro	Ile	Gly	Pro	Pro	Gly	Pro	Arg	Gly	Arg	Thr	Gly	Asp	Ala	Gly
						165					170					175	
35		Pro	Val	Glv	Pro	Pro	Glv	Pro	Pro	Glv	Pro	Pro	Gly	Pro	Pro	Gly	Pro
					180		•			185			-		190	-	
;																	
40.		Pro	Ser		Gly	Phe	qaA	Phe		Phe	Leu	Pro	Gln		Pro	Gln	Glu
				195					200					205			
45		Lys	Ala	His	Asp	Gly	Gly	Arg	Tyr	Tyr	Arg	Ala					
			210	ı				215									
50:	(2)	INFO	RMAT	NOI	FOR	SEQ	ID N	0:45	:								
					_												
,		(i)			E CH												
55			(A)	., DE	MGIU	. 02	, Da	ae b	411 B								

(B) TYPE: nucleic acid(C) STRANDEDNESS: single

		(D) TOPOLOGY	: linear				
10 <sup>.</sup>	(ii) M	OLECULE TYPE	E: cDNA				
15	(xi) SI	EQUENCE DESC	CRIPTION: SI	EQ ID NO:45	:		
	ATGGGCTCTC	CGGGTGTTAA	CGGCGCCCCT	GGTGAAGCGG	GCCGCGACGG	CAATCCGGGC	60
20 <sup>.</sup>	AACGATGGTC	CGCCGGGTCG	TGATGGTCAG	CCGGGTCACA	AAGGTGAGCG	TGGCTACCCG	120
	GGTAACATCG	GTCCGGTTGG	TGCGGCCGGC	GCTCCGGGTC	CGCACGGTCC	GGTAGGCCCA	180
25	GCCGGCAAAC	ACGGTAACCG	TGGTGAAACG	GGTCCGTCCG	GTCCGGTAGG	TCCGGCGGGT	240
<b>.</b>	GCTGTTGGTC	CACGCGGCCC	GTCCGGCCCG	CAGGGTATTC	GCGGTGACAA	AGGCGAACCG	300
30 <sup>.</sup>	GGCGAAAAAG	GTCCGCGTGG	TCTGCCGGGC	CTTAAGGGCC	ACAACGGTCT	GCAAGGTCTG	360
<b>35</b>	CCGGGTATCG	CGGGTCACCA	CGGTGATCAG	GGTGCTCCGG	GTTCCGTTGG	TCCGGCCGGT	420
	CCGCGTGGCC	CGGCTGGTCC	GTCTGGTCCG	GCCGGTAAAG	ACGGCCGTAC	GGGCCACCCG	480
40 <sup>.</sup>	GGTACGGTGG	GTCCGGCCGG	CATTCGCGGT	CCGCAAGGTC	ACCAGGGTCC	GGCGGGTCCG	540
45	CCGGGTCCGC	CGGGTCCGCC	GGGTCCGCCG	GGTGTTAGCG	GTGGCGGTTA	TGATTTTGGT	600
	TATGACGGTG	ATTTCTATCG	TGCGTAA				627
50 <sup>-</sup>							
, 55							

ι. εν

5	(2)	INFOR	MATI	ON F	OR S	EQ I	D NC	: 46 :									
		(i)	_		СНА												
					IGTH :				cids	I							
10			(B)	TYP	E: a	mino	aci	.d									
			(C)	STR	LANDE	DNES	SS: 8	ingl	. <b>e</b>								
			(D)	TOP	OLOG	SY: t	ınkno	wn									
15																	
		(ii)	MOLE	CULE	TYE	E: p	epti	de									
20		(xi)	SEQU	JENCE	E DES	CRIE	PTION	1: SE	EQ II	ONO:	46:						
		Mot	Gly	Pro	Dro	Glv	T. <b>A</b> 11	בומ	G] v	Pro	Dro	Glv	Glu	Ser	Gly	Ara	Gl 11
			Gry	<b>P10</b>		5	Deu	AIG	GLY	110	10	UL,		JU2		15	<b>924</b>
25		1				,					10					13	
										_			_		_	_	
		Gly	Ala	Pro	-	Ala	Glu	Gly	Ser	Pro	GIA	Arg	Asp	GIA	Ser	Pro	Gly
					20					25					30		
30																	
·		Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Ala
				35					40					45			
35																	
		Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Pro	Val	Gly	Pro	Ala	Gly	Lys	Ser
			50					55					60				
							,										
40		Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro	Val	Gly
		65	•	Ī	-		70	•			-	75		•			80
		0.5															
45		Pro	a l a	ഭിഴ	Δla	Δνα	Glv	Pro	Δla	Gl v	Pro	Gln	Glv	Pro	Arg	Glv	Δan
		PIO	AIG	Gry	7.2.0	85	017	110	714	GLY		<b>G111</b>	017		AL 9	95	nop
						03					90					73	
		_				~.				_	_			_			
50		Lys	Gly	Glu		GLY	Glu	Gln	Gly	_	Arg	Gly	Ile	ГÀ8	Gly	His	Arg
					100					105					110		

	Gly	Phe	Ser	Gly	Leu	Gln	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Ser	Pro	Gly	
			115					120					125				
		<b>~</b> 3	<b>~1</b>	<b>D</b>	<b>9</b>	<b>01</b>	• • • •	<b>0</b>	<b>01</b>		71.	<i>a</i> 1	Dwa	7	<b>01</b>	<b>5</b>	
0	Glu		GIÀ	Pro	ser	GIY		ser	GIA	Pro	Ala	140	Pro	Arg	Gly	Pro	
•		130					135					140					
	Pro	Gly	Ser	Ala	Gly	Ala	Pro	Gly	Lys	Asp	Gly	Leu	Asn	Gly	Leu	Pro	
5	145				_	150		_			155					160	
	Gly	Pro	Ile	Gly	Pro	Pro	Gly	Pro	Arg	Gly	Arg	Thr	Gly	Asp	Ala	Gly	
20					165					170					175		
	Pro	Val	Gly		Pro	Gly	Pro	Pro	-	Pro	Pro	Gly	Pro		Gly	Pro	
?5				180					185					190			
	7-1-		21-	<b>~1</b>	Dho	n a m	Dho	Com	Dha	T ON	Pro	Gln	Pro	Pro	Gln	Glu	
	Pro	Ser	195	GIY	PHE	web	PHE	200	FIIE	nea	FIU	GIII	205	FIO	GIII	Giu	
30			195					200					203				
	Lys	Ala	His	Asp	Gly	Gly	Arg	Tyr	Tyr	Arg	Ala						
		210					215										
35																	
	(2) INFO	RMAT:	ION	FOR :	SEQ	ID N	0:47	:									
		a=0	· • • • • • • • • • • • • • • • • • • •	- AI	3 D 3 C	man t	OMTO	<b>c</b> .									
10	(1)	SEQ!				bas											
•						eic											
		•	-			SS:											
15		(D	) TO	POLO	GY:	line	ar										
	(ii)	MOL	ECUL	E TY	PE:	CDNA											
50																	

<sup>3</sup> 5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:	
	GGAATTCTCC CATGGGCCCG CCGGGTCTGG CGGGCCCTCC GGGTGAAAGC GGTCGTGAAG	60
<sup>2</sup> 10	GCGCGCCGGG TGCCGAAGGC AGCCCAGGCC GCGAC	95
	(2) INFORMATION FOR SEQ ID NO:48:	
15	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 97 hase pairs	
	(B) TYPE: nucleic acid	
20	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	-
25	(ii) MOLECULE TYPE: cDNA	
÷	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:	
30	CTTCCGTCGG GTCCGGCGCT GCCATCGGGC CCCCGGTTTC CCCTAGCACC ACTTTGGCCG	60
35	GGCCGCCCGG GGGGCCCACG TGGCATTATT CGAACCC	97
	(2) INFORMATION FOR SEQ ID NO:49:	
40	(i) SEQUENCE CHARACTERISTICS:	
40	(A) LENGTH: 91 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
45	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
50		

147

. 55

5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:	
J	GGAATTCGGT GCACCGGGCG CGCCGGGTGC CCCAGGCCCG GTGGGCCCGG CGGGCAAAAG	60
10	CGGTGATCGT GGCGAGACCG GTCCGGCGGG C	91
Ċ.	(2) INFORMATION FOR SEQ ID NO:50:	
20	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 91 base pairs  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single	
25	(ii) MOLECULE TYPE: cDNA	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:	
*	CTCTGGCCAG GCCGCCCGGG CCGGCCAGGC CACCCGGGTC GCCCGCGGGC ACCGGGCCGG	60
35	CCAGGCGTCC CGGGCGCCAT TATTCGAACC C	91
40 <sup>-</sup>	Claims	
<b>45</b>	<ol> <li>A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof capable of providing a gregate in a cell which does not ordinarily hydroxylate proline comprising</li> <li>providing a nucleic acid sequence encoding the EMP or fragment thereof which has been optimize pression in the cell by substitution of codons preferred by the cell for naturally occurring codons not p by the cell;</li> </ol>	d for ex-
50	incorporating the nucleic acid sequence into the cell; providing hypertonic growth media containing at least one amino acid selected from the group cons trans-4-hydroxyproline and 3-hydroxyproline; and contacting the cell with the growth media wherein the at least one amino acid is assimilated into the incorporated into the EMP or fragment thereof.	_
55 <sub>.</sub>	2. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 1 who EMP is selected from the group consisting of human collagen, fibrinogen, fibronectin and collagen-like p	

3. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 1 or 2, wherein

the cell is a prokaryote.

5

10

15

20

25.

30

35

40

45

- 4. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 3, wherein the prokaryote is E. coli.
- A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claims 2 4, wherein the human collagen is Type I (α1).
- A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 5, wherein the nucleic acid encoding human collagen Type I (α1) includes the sequence shown in SEQ.ID.NO.19.
  - A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claim 2 to 4, wherein the human collagen is Type I (α2).
- 8. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 7, wherein the nucleic acid encoding human collagen Type I (α2)= includes the sequence shown in SEQ.ID.NO.31.
- 9. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claims 1 to 8, wherein the nucleic acid encoding the EMP includes the sequence shown in SEQ.ID.NO. 43.
- 10. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claims 1 to 8, wherein the nucleic acid encoding the EMP includes the sequence shown in SEQ.ID.NO. 46.
- 11. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claims 1 to 10, wherein the nucleic acid sequence includes nucleic acid encoding a physiologically active peptide.
  - 12. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 11, wherein the physiologically active peptide is selected from the group consisting of bone morphogenic protein, transforming growth factor-β and decorin.
  - 13. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claims 1 to 4, wherein the EMP or fragment thereof is a collagen-like peptide.
  - 14. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 13, wherein the EMP or fragment thereof includes the amino acid sequence depicted in SEQ.ID.NO. 4.
  - 15. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 13, wherein the EMP includes the amino acid sequence depicted in SEQ.ID.NO.40.
- 16. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 1, wherein the EMP includes the amino acid sequence depicted in SEQ.ID.NO. 44.
- 17. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 1, wherein the EMP is a collagen fragment including the amino acid sequence depicted in SEQ.ID.NO. 26.
- 18. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 1, wherein the EMP is a collagen fragment including the amino acid sequence depicted in SEQ.ID.NO. 46.
- 19. Nucleic acid encoding a chimeric protein comprising a domain from a physiologically active peptide and a domain from an Extracellular Matrix Protein (EMP) which is capable of providing a self-aggregate.
- 20. Nucleic acid encoding a chimeric protein according to claim 19, wherein said EMP is selected from the group consisting of human collagen, fibrinogen, fibronectin and collagen-like peptide.
- 21. Nucleic acid encoding a chimeric protein according to claim 19 or 20 wherein said domain from a physiologically active peptide is selected from the group consisting of bone morphogenic protein, transforming growth factor β and decorin.

- 22. Nucleic acid encoding a chimeric protein according to any of claims 19 21, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.6.
- 23. Nucleic acid encoding a chimeric protein according to any of claims 19 21, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.8.
- 24. Nucleic acid encoding a chimeric protein according to any of claims 19 21, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.11.
- 25. Nucleic acid encoding a chimeric protein according to any of claims 19 21, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.10.
  - 26. A cloning vector comprising nucleic acid according to any of claims 19 21.
- 27. A cloning vector according to claim 26 wherein said cloning vector is selected from the group consisting of plasmid, phage, cosmid and artificial chromosome.
  - 28. A cell transformed by a vector according to claim 26 or 27.

5

25

40

45

- 20 29. A chimeric protein comprising a domain from a physiologically active peptide and a domain from an Extracellular Matrix Protein (EMP) which is capable of providing a self-aggregate.
  - **30.** A chimeric protein according to claim 29 wherein said EMP is selected from the group consisting of human collagen, fibrinogen, fibronectin and collagen-like peptide.
  - 31. A chimeric protein according to claim 29 or 30 wherein said domain from a physiologically active peptide is selected from the group consisting of bone morphogenic protein, transforming growth factor β and decorin.
- **32.** A chimeric protein according to any of claims 29 31, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.6.
  - 33. A chimeric protein according to any of claims 29 31, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.8.
- 35 **34.** A chimeric protein according to any of claims 29 31, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.10.
  - 35. A chimeric protein according to any of claims 29 31, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.11.
  - 36. Human collagen or fragment thereof produced by a prokaryotic cell, the human collagen or fragment thereof being capable of providing a self-aggregate.
  - 37. Human collagen or fragment thereof produced by a prokaryotic cell according to claim 36 wherein the human collagen or fragment thereof is encoded for by nucleic acid having the sequence shown in SEQ.ID.NO.19.
    - 38. Human collagen or fragment thereof produced by a prokaryotic cell according to claim 36 wherein the human collagen or fragment thereof is encoded for by nucleic acid having the sequence shown in SEQ.ID.NO.39.
- 39. Human collagen or fragment thereof produced by a prokaryotic cell according to claim 36 wherein the human collagen or fragment thereof is encoded for by nucleic acid having the sequence shown in SEQ.ID.NO.43.
  - **40.** Human collagen or fragment thereof produced by a prokaryotic cell according to claim 36 wherein the human collagen or fragment thereof is encoded for by nucleic acid having the sequence shown in SEQ.ID.NO.45.
  - 41. Human collagen or fragment thereof produced by a prokaryotic cell according to claim 36 wherein the collagen or fragment thereof is encoded for by nucleic acid having the sequence shown in SEQ.ID.NO.31.

- 42. Nucleic acid comprising the sequence shown in SEQ.ID.NO. 19.
- 43. Nucleic acid comprising the sequence shown in SEQ.ID.NO. 31.
- 44. Nucleic acid comprising the sequence shown in SEQ.ID.NO. 43.

5

10

15

20

25

30

35

40

45

50

- 45. Nucleic acid comprising the sequence shown in SEQ.ID.NO. 45.
- 46. Nucleic acid encoding a human Extracellular Matrix Protein (EMP) or fragment thereof wherein the codon usage in the nucleic acid sequence reflects preferred codon usage in a prokaryotic cell.
- 47. Nucleic acid according to claim 46 wherein the prokaryotic cell is *E. coli*.
- 48. Nucleic acid according to claim 43 wherein the EMP is selected from the group consisting of collagen, fibrinogen, fibronectin and collagen-like peptide.

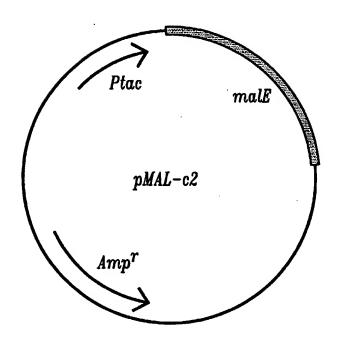
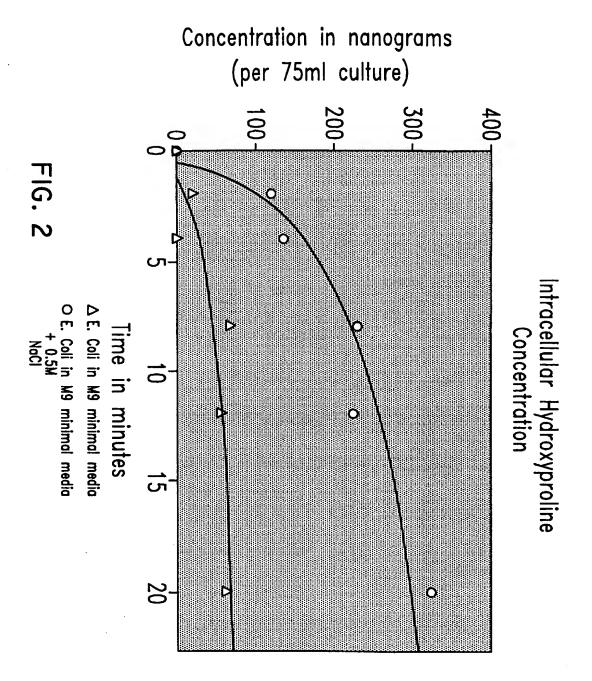


FIG. I



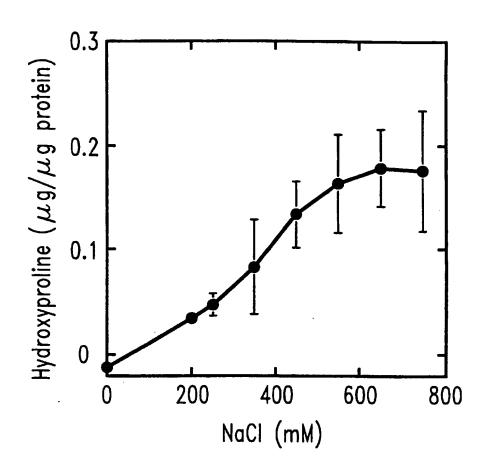


FIG. 2A

5'- CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC TEGECECATE SETECETETE GIVETEGTES TETECETEGE CECECTEGTS CACCTGGTCC CCAAGGCTTC CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT GGTGAGCGTG GCCTCCTGG GCCTCAGGGT GCTCGAGGAT TGCCCGGAAC AGCTGGCCTC CCTGGAATGA AGGGACACAG AGGTTTCAGT GGTTTGGATG GTGCCAAGGG AGATGCTGGT CCTGCTGGTC CTAAGGGTGA GCCTGGCAGC CCTGGTGAAA ATGGAGCTCC TGGTCAGATG GGCCCCGTG GCCTGCCTGG TGAGAGAGGT CGCCCTGGAG CCCCTGGCCC TGCTGGTGCT CGTGGAAATG ATGGTGCTAC TEGTECTECC GESCUCCTG GTOCCACCGG CCCGGCTGGT CCTCCTGGCT TCCCTGGTGC TGTTGGTGCT AAGGGTCAAG CTGGTCCCCA AGGGCCCCGA GECTCTGAAG GTCCCCAGGG TGTGCGTGGT GAGCCTGGCC CCCCTGGCCC TECTEGTECT CCTEGCCCTG CTGCAAACCC TGGTGCTGAT GGACAGCCTG GTGCTAAAGG TGCCAATGGT GCTCCTGGTA TTGCTGGTGC TCCTGGCTTC CCTEGTECCC GAGECCCCTC TEGACCCCAG GECCCCGCG GCCCTCCTGG TCCCAAGGGT AACAGCGGTG AACCTGGTGC TCCTGGCAGC AAACGAGACA CTCGTCCTAA GGCAGACCCT GCCCCTGTTG GTGTTCAAGG ACCCCCTGGC CCTCCTGGAG AGGAAGGAAA GCGAGGAGCT CGAGGTGAAC CCGGACCCAC TEGECETECCE GEACCECETG GEGAGEGTEG TEGACETEGT ACCEGTEGTT TCCCTGGCGC AGATGGTGTT GCTGGTCCCA AGGGTCCCGC TCGTGAACGT CGTTCTCCTG GCCCCGCTGG CCCCAAAGGA TCTCCTGGTG AAGCTGGTCG TCCCGGTGAA GCTGGTCTGC CTGGTGCCAA GGGTCTGACT GGAAGCCCTG CCACCCTGG TCCTGATGGC AAAACTCGCC CCCCTGGTCC CGCCGGTCAA CATGGTCGCC COGGACCCCC AGGCCCACCT GGTGCCCGTG GTCAGGCTGG TGTGATGGGA TTCCCTGGAC CTAAAGGTGC TGCTGGAGAG CCCGGCAAGG CTGGAGAGGG AGGTGTTCCC GGACCCCCTG GCGCTGTCGG TCCTGCTGGC AAAGATGGAG AGGCTGGAGC TCAGGGACCC CCTGGCCCTG CTGGTCCCGC TGGCGAGAGA GGTGAACAAG GCCCTGCTGG CTCCCCCGGA TTCCAGGGTC TCCCTGGTCC TGCTGGTCCT CCAGGTGAAG CAGGCAAACC TGGTGAACAG GGTGTTCCTG CAGACCTTGG CGCCCCTGCC CCCTCTGGAG CAAGAGGCCGA CAGAGGTTTC CCTGGCGAGC GTGGTGTCCA AGGTCCCCCT GGTCCTGCTG CACCCCGAGG GGCCAACGGT GCTCCCGGCA ACGATGGTGC TAAGGGTGAT CCTGGTGCCC CTGGAGCTCC CGGTAGCCAG GGCGCCCCTG GCCTTCAGGG AATGCCTGGT GAACGTGGTG CACCTGGTCT TCCAGGGCCCT AAGGGTGACA CAGGICATIC TEGTECCAAA GGTGCTGATG GCTCTCCTGG CAAAGATGGC

FIG. 3A

CICCCICCIC LCYCCCCCCC CYILCELCCI, CCLCCCCLL CLCCLCCCC
TGGTGACAAG GGTGAAAGTG GTOCCAGGGG CCCTGCTGGT CCCACTGGAG
CTCGTCGTCC CCCCCGAGAC CGTCGTGAGC CTCGTCCCCC CCCCCCTCCT
GECTTTECTG GCCCCCCTGG TGCTGACGGC CAACCTGGTG CTAAAGGCGA
ACCTGGTGAT GCTGGTGCCA AAGGCGATGC TGGTCCCCCT GGGCCTGCCG
GACCCCCTCG ACCCCCTCGC CCCATTGGTA ATGTTCGTCC TCCTGCAGCC
AAAGGTGCTC GCCGCAGCGC TGGTCCCCCT GGTGCTACTG GTTTCCCTGG
TECTECTESC CEASTCESTC CTCCTGGCCC CTCTGGAAAT GCTGGACCCC
CTGGCCCTCC TGGTCCTGCT GGCAAAGAAG GCGGCAAAGG TCCCCGTGGT
GAGACTGGCC CTGCTGGACG TCCTGGTGAA GTTGGTCCCC CTGGTCCCCC
TECCETECT GEOGRAPAG CATECETES TECTEATEST CETECTESTS
CTCCTGGTAC TCCCGGGCCT CAAGGTATTG CTGCACAGCG TGGTGTGGTC
GCCCTGCCTG GTCAGAGAGG AGAGAGAGGC TTCCCTGGTC TTCCTGGCCC
CTCTCGTGAA CCTGGCAAAC AAGGTCCCTC TGGAGCAAGT GGTGAACGTG
CICCCCCCC TCCCATGGGC CCCCCTGGAT TGGCTGGACC CCCTGGTGAA
TCTGGACGTG ACCGGCCTCC TGCTGCCGAA GGTTCCCCTG GACGAGACGG
TTCTCCTGGC GCCAAGGGTG ACCGTGGTGA GACCGGCCCC GCTGGACCCC
CTEGTECTCC TEGTECTCCT GETECCCCTG GCCCCGTTGG CCCTGCTGGC
AACAGTCGTG ATCGTCGTCA CACTCGTCCT CCTCGTCCCGT
CCCCCCCCCT CCCCCCCTG CCCCCCCCG ACCCCAACCC CCCCCTGGTG
ACAAGGGTGA GACAGGCGAA CAGGGCGACA GAGGCATAAA GGGTCACCGT
GCCTTCTCTG GCCTCCAGGG TCCCCCTGGC CCTCCTGGCT CTCCTGGTGA
ACAAGGTCCC TCTGGAGCCT CTGGTCCTGC TGGTCCCCGA GGTCCCCCTG
SCTCTCCTCG TCCTCCTCSC AAACATCCAC TCAACCGTCT CCCTGGCCCC
ATTECCCCC CTCGTCCTCG CCGTCCCACT CGTCATCCTG GTCCTGTTCG
TCCCCCCGGC CCTCCTGGAC CTCCTGGTCC CCCTGGTCCT CCCAGGGCTG
STITICGACTT CAGCTTCCTC CCCCAGCCAC CTCAAGAGAA GGCTCACGAT
GETGECOGGT ACTACCGGGC T-3

FIG. 3B

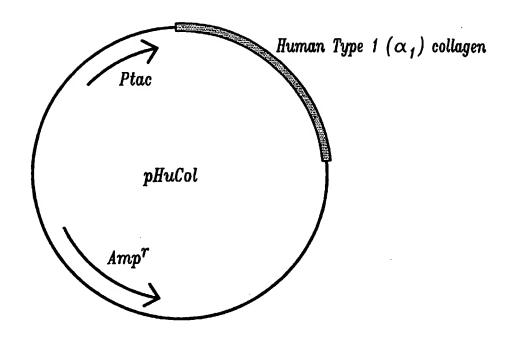


FIG. 4

5'- CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC
TGGCCCCATG GGTCCCTCTG GTCCTCGTGG TCTCCCTGGC CCCCCTGGTG
CACCTGGTCC CCAAGGCTTC CAAGGTCCCC CTGGTGACCC TGCCGAGCCT
GGAGCTTCAG GTCCCATGGG TCCCCGAGGT CCCCCAGGTC CCCCTGGAAA
GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT-3'

FIG. 5

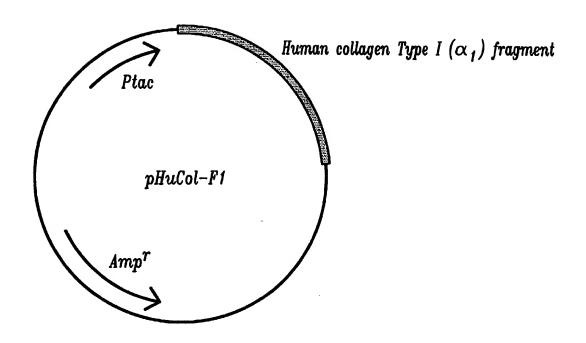


FIG. 6

GGA TCC ATG GGG CTC GCT GGC CCA CCG GGC GAA CCG GGT CCG CCA GGC CCG AAA GGT CCG CGT GGC GAT AGC GGG CTC CCG GGC GAT TCC TAA TGG ATC C

Gly-Leu-Ala-Gly-Pro-Pro-Gly-Glu-Pro-Gly-Pro-Pro-Gly-Pro-Lys-Gly-Pro-Arg-Gly-Asp-Ser

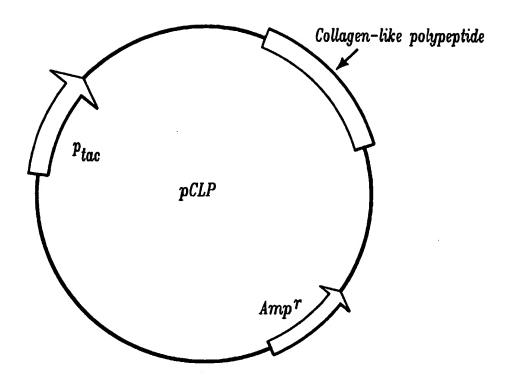


FIG. 9

5'-	*********	GGAAGAAGAA	TAAGAACTGC	CGGCGCCACT	CGCTCTATGT
	GGACTTCAGC	GATGTGGGCT	GGAATGACTG	GATTGTGGCC	CCACCAGGCT
	ACCAGGCCTT	CTACTGCCAT	GGGGACTGCC	CCTTTCCACT	GGCTGACCAC
	CTCAACTCAA	CCAACCATGC	CATTGTGCAG	ACCCTGGTCA	ATTCTGTCAA
	TTCCAGTATC	CCCAAAGCCT	GTTGTGTGCC	CACTGAACTG	AGTGCCATCT
	CCATGCTGTA	CCTGGATGAG	TATGATAAGG	TGGTACTGAA	AAATTATCAG
		TAGAGGGATG			

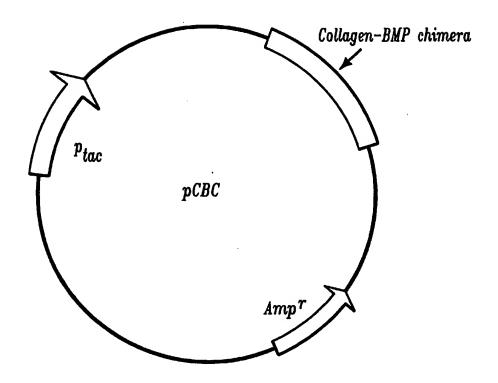


FIG. II

# Mole percent of MBP

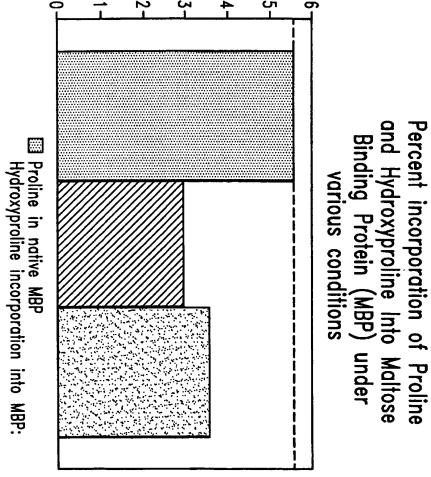


FIG. 12

🖾 Grown in hypertonic media

☐ Grown in isotonic media

10 QLSYGYDEKS	20 TGGISVPGPM	30 GPSGPRGLPG	40 PPGAPGPQGF	SO OGPPGEPGEP	60 GASGPMGPRG
70 PPGPPGKNGD	80 DGEAGXPGRP	90 GERGPPGPQG	100 Arglpgtagl	110 PGMKGHRGFS	120 GLDGAXGDAG
130 PAGPKGEPGS	140 PGENGAPGQM	150 GPRGLPGERG	160 RPGAPGPAGA	170 RONDGATGAA	180 GPPGPTGPAG
190 PPGFPGAVGA	200 KGEAGPQGPR		220 EPGPPGPAGA		240 GQPGAKGANG
250 APGIAGAPGF	260 PGARGPSGPQ	270 GPGGPPGPKG	280 NSGEPGAPGS	290 KGDTGAKGEP	300 GPVGVQGPPG
310 PAGEEGKRGA	320 RGEPGPTGLP	330 GPPGERGGPG	340 SRGFPGADGV	350 AGPXGPAGER	360 GSPGPAGPKG
370 SPGEAGRPGE					420 GARGQAGVHG
430 FPGPKGAAGE		450 GPPGAVGPAG	460 KDGEAGAQGP	470 PGPAGPAGER	480 GEQGPAGSPG
490 FQGLPGPAGP	500 PGEAGKPGEQ	510 GVPGDLGAPG	520 PSGARGERGF	530 PGERGVQGPP	540 GPAGPRGANG
550	560	570	580	590	600 GADGSPGKDG
610 VRGLTGPIGP		630 GESGPSGPAG	640 PTGARGAPGD	650 RGEPGPPGPA	660 GFAGPPGADG
670	680	690	700		720
730	740	750	760	770 VGPPGPPGPA	780
790	800	810	820	830 PGKQGPSGAS	840
850	860	870	880	890 AGPPGAXGAX	900
	920	. 930	940	950	960
970 PPGSPGEQGP	.980	990	1000	1010	1020
1030 PPGPPGPPGP	1040	1050	1060	1070	1080
1090 VGWWWIVAP	1100	1110	1120	1130	1140
1150	1160 DKVVLKNYQE	1170	1180	1190	

FIG. 13

10	20	30	40	. 50	60
gggaaggatt	tccatttccC	AGCTGTCTTA	TGGCTNTGAT	GAGAAATCAA	CCGGAGGAAT
70	80	90	100	. 110	120 CCCCTGGTGC
130	140	150	160	170	180
ACCTGGTCCC	CAAGGCTTCC	AAGGTCCCCC	TGGTGNGCCT	GGCGAGCCTG	GAGCTTCAGG
190	200	210	220	230	240
TCCCATGGGT	CCCCGAGGTC	CCCCAGGTCC	CCCTGILAAAG	AATGGAGATG	ATGGGGAAGC
250	260	270	280	290	300
TGGAAAACCT	GGTCGTCCTG	GTGAGCGTGG	GCCTC:TTGGG	CCTCAGGGTG	CTCGAGGATT
310	320	330	340	350	360
GCCCGGAACA	GCTGGCCTCC	CTGGAATGAA	GGGAC.\CAGA	GGTTTCAGTG	GTTTGGATGG
370	380	390	400	410	420
TGCCAAGGGA	GATGCTGGTC	CTGCTGGTCC	TAAGGTTGAG	CCTGGCAGCC	CTGGTGAAAA
430	.440	450	460	470	480
TGGAGCTCCT	GGTCAGATGG	GCCCCCGTGG	CCTGCCTGGT	GAGAGAGGTC	GCCCTGGAGC
490	500	510	520	530	540
CCCTGGCCCT	GCTGGTGCTC	GTGGAAATGA	TGGTG:TACT	GGTGCTGCCG	GGCCCCCTGG
550	560	570	580	590	600
TCCCACCGGC	CCCGCTGGTC	CTCCTGGCTT	CCCTGTTGCT	GTTGGTGCTA	AGGGTGAAGC
610	620	630	640	650	. 660
TGGTCCCCAA	GGGCCCCGAG	GCTCTGAAGG	TCCCC.4GGGT	GTGCGTGGTG	AGCCTGGCCC
670	680	690	700	710	720
CCCTGGCCCT	GCTGGTGCTG	CTGGCCCTGC	TGGAŁICCCT	GGTGCTGATG	GACAGCCTGG
730	740	750	760	770	780
TGCTAAAGGT	GCCAATGGTG	CTCCTGGTAT	TGCTOJTGCT	CCTGGCTTCC	CTGGTGCCCG
790 AGGCCCCTCT	800 GGACCCCAGG		820 CCCTCCTGGT		
ACCTGGTGCT	CCTGGCAGCA	AAGGAGACAC	TGGTGTTAAG	GGAGAGCCTG	
910	920	930	940	950	960
TGTTCAAGQA	CCCCCTGGCC	CTGCTGGAGA	GGAAG:2AAAG	CGAGGAGCTC	GAGGTGAACC
	980 GGCCTGCCCG	990 GACCCCCTGG			1020 GCCGTGGTTT
1030	1040	1050	1060	1070	1080
CCCTGGCGCA	GATGGTGTTG	CTGGTCCCAA	GGGTC:CGCT	COTGAACGTG	GTTCTCCTGG
1090	1100	1110	1120	1130	1140
CCCCGCTGGC	CCCAAAGGAT	CTCCTGGTGA	AGCTG:TCGT	CCCGGTGAAG	CTGGTCTGCC
1150 TGGTGCCAAG			1180 CAGCCCTGGT		1200 AAACTGGCCC
1210	1220	1230	CGGACUCCCA	GGCCCACCTG	1260
CCCTGGTCCC	GCCGGTCAAG	ATGGTCGCCC	FIG. 14		GTGCCCGTGG

1270 TCAGGCTGGT	1280 GTGATGGGAT	1290 TCCCTGGACC	1300 TAAAGGTGCT	· 1310 GCTGGAGAGC	CCGGCAAGGC
1220	1340	1350	1360	1370	1380
	GGTGTTCCCG				
1390 GGCTGGAGCT	1400 CAGGGACCCC	1410 CTGGCCCTGC	TGGTC::CGCT	GGCGAGAGAG	GTGAACAAGG
1450	1460 TCCCCCGGAT	1470	1480	1490 GCTGGTCCTC	1500 CAGGTGAAGC
					1560
1510 AGGCAAACCT	GGTGAACAGG	GIGIICCIGG	AGACC'ITGGC	GUCCLAGGCC	CCTCTGGAGC
1570	1580	1590 CTGGCGAGCG	1600 TGGTG'IGCAA	1610 GGTCCCCCTG	1620 GTCCTGCTGG
1630	1640	1650	1660	1670	1680
ACCCCGAGGG	GCCAACGGTG	CTCCCGGCAA	CGATGGTGCT	AAGGGTGATG	CTGGTGCCCC
1690 TOGAGCTCCC	1700 GGTAGCCAGG	1710 GCGCCCTGG	1720 CCTTC\GGGA	1730 ATGCCTGGTG	1740 AACGTGGTGC
1750	1760	1770	1780	1790	1800
				•	GTGCTGATGG
1810 CTCTCCTGGC	1820 AAAGATGGCG	TCCGTGGTCT	1840 GACÇG3CCCC	1850 ATTGGTCCTC	1860 CTGGCCCTGC
1870	1880 GGTGACAAGG	1890	1900	1910	1920
	1940				
					CCTTTCCTCC
1990 CCCCCCTGGT	2000 GCTGACGGCC				2040 CTGGTGCCAA
2050	2060	2070	2080	2090	2100
	CGTCCCCCTG	•		٠.	
TGTTGGTGCT	2120 CCTGGAGCCA	AAGGTGCTCG	CGGCAGCGCT	GGTCCCCCTG	2160 GTGCTACTGG
2170	2180	2190	2200	2210	2220
•	GCTGCTGGCC			;	•
					2280 - AGACTGGCCC
					2340
					CCGAGAXAGG
					2400 AAGGTATTGC
				2450 GAGAGAGGCT	2460 TCCCTGGTCT
	2480 TCTGGTGAAC				2520 GTGAACGTGG

FIG. 14B

2530 TCCCCCCGGT	2540 CCCATGGGCC	2550 CCCCTGGATT	2560 GGCTGGACCC	· 2570	2580 CTGGACGTGA
2590 GGGGGCTCCT	2600 GCTGCCGAAG	2610 GTTCCCCTGG	2620 ACGAGACGGT	2630 TCTCCTGGCG	2640 CCAAGGGTGA
2650 CCGTGGTGAG	2660 200000000	2670 CTGGACCCCC	2680 TGGTGCTCNT	2690 GGTGCTCNTG	2700 GTGCCCCTGG
2710	2720	2730	2740	2750	2760
CCCCGTTGGC	CCTGCTGGCA	AGAGTGGTGA	TCGTGCTGAG	ACTGGTCCTG	CTGGTCCCGC
2770	2780	2790	2800	2810	2820
					CCCGTGGTGA
2830 CAAGGGTGAG	2840 ACAGGCGAAC	2850 AGGGCGACAG	2860 AGGCATAAAG	2870 GGTCACCGTG	2880 GCTTCTCTGG
2890 CCTCCAGGGT	2900	2910 CTCCTGGCTC	2920	2930	2940 CTGGAGCCTC
					3000
TOGTCCTGCT	CCTCCCCCAC	CTCCCCTCC	CTCTGCTGGT	GCTCCTGGCA	AAGATGGACT
3010	3020	3030	3040	3050	3060
CAACGGTCTC	CCTGGCCCCA	TTGGGCCCCC	TOGTCCTCGC	CÉTCCCYCIC	GTGATGCTGG
3070	3080	3090	3100	3110	3120
					CCAGCGCTGG
3130	3140	3150	3160	3170	3180
					GTGGCCGCTA
3190	3200	3210	3220	3230	3240 GCCACTCGCT
3250	3260	3270	3280	3290	3300
					CAGGCTACCA
GGCCTTCTAC	3320	3330	3340	3350	3360 .
	TGCCATGGG				
3370 CCATGCCATT	3380	3390	3400	3410	3420
	GTGCAGACCC				
3430	3440	3450	3460	3470	3480
TGTGCCCACT		•			
3490 T44447T3A	3500	3510	3520	3530	3540
ACTGAAAAAT	יאי כאטטאטא	1001MOTARY	COCATOTOGG	TUCCGCTAAa	agctt

FIG. 14C

60 Gasgphgprg	50 QGPPGEPGEP	40 PPGAPGPQGF	30 GPSGPRGLPG	20 TGGISVPGPM	10 QLSYGYDEKS
120	110	100	00		
180 GPPGPTGPAG	170 RGNDGATGAA	160 RPGAPGPAGA	150 GPRGLPGERG	140 PGENGAPGQM	130 PAGPKGEPGS
240 GQPGAXGANG	230 AGPAGNPGAD	220 EPGPPGPAGA	210 GSEGPQGVRG	200 KGEAGPQGPR	
300 GPVGVQGPPG	290 KGDTGAKGEP	280 NSGEPGAPGS	270 GPGGPPGPKG	260 PGARGPSGPO	250 APGIAGAPGF
360 GSPGPAGPXG	350	340	330	, , , , , ,	
420 Gargoagviig	410	.400	390	380	220
480 GEQGPAGSPG	470	460	450	440	. 430
540 GPAGPRGANG	530	520	510	500	490
600 GADGSPGKDG	590	580	570	\$60	550
660 GFAGPPGADG	650	640	. 630	620	610
720 GATGFPGAAG	710	700	690	680	670
780 GEKGSPGADG	770	760	750	740	730
840 GERGPPGPMG	830	820	810	. 800	790
900 GAPGPVGPAG	890	880	870	860	850
960 GFSGLQGPPG	· 950	940	930	920	910
1020 GDAGPVGPPG	. 1010	1000	990	980	970
1080 EXHCCVRQLY	1070	1060	1050	1040	1030
_	1130	1120	1110	1100	1090
	1190	1180	1170	1160	1150

FIG. 15

10 gggaaggatt	20 tccatttccC	30 AGCTGTCTTA	40 TGGCTATGAT	50 GAGAAATCAA	60 CCGGAGGAAT
	90	00	100	110	120
		150	160	170	180
	200	210	. 220	230	240 ATGGGGAAGC
750	260	270	280	290	300 CTCGAGGATT
210	320	330	340	350	360 GTTTGGATGG
370	380	390	400	410	420
430	440	450	460	470	CTGGTGAAAA 480
TGGAGCTCCT	GGTCAGATGG	eccicccesee	CCTGCCTGGT	GAGAGAGGTC	GCCCTGGAGC 540
CCCTGGCCCT	GCTGGTGCTC	GTGGAAATGA	TGGTG/TACT	GGTGCTGCCG	GCCCCCTGG
TCCCACCGGC	CCCCCTGGTC	CTCCTGGCTT	CCCTC/ITGCT	CTTGCTGCTA	AGGGTGAAGC
610 TGGTCCCCAA	620 GGGCCCCGAG	. 630 GCTCTGAAGG	640 TCCCCAGGGT	650 GTGCGTGGTG	660 AGCCTGGCCC
670 CCCTGGCCCT	680 GCTGGTGCTG	690 CTGGCCCTGC	700 TGGAAACCCT	710 GGTGCTGATG	720 GACACCCTGG
730 TGCTAAAGGT	740 GCCAATGGTG	750 CTCCTGGTAT	760 TGCTGGTGCT	770 CCTGGCTTCC	780 CTGGTGCCCG
790	800	810	820	830	840 ACAGCGGTGA
850 ACCTGGTGCT	860 CCTGGCAGCA	870 AAGGAGACAC	880 TGGTGCTAAG	890 GGAGAGCCTG	900 GCCCTGTTGG
910	920	930	940	950	960 GAGGTGAACC
970	.980	990	1000	. 1010	1020
. 1030	1040	1050	1060	. 1070	GCCGTGGTTT 1080
	GATGGTGTTG 1100				GTTCTCCTGG
CCCCGCTGGC	CCCAAAGGAT	CTCCTGGTGA	ACCTGGTCGT	CCCGGTGAAG	CTGGTCTGCC
TGGTGCCAAG	GGTCTGACTG	GAAGCCCTGG	CAGCCCTGGT	CCTGATGGCA	AAACTGGCCC
	1220 GCCGGTCAAG		1240 CGGACCCCCA	1250 GGCCCACCTG	1260 GTGCCCGTGG

FIG. 16A

1270	1280	1290	1300	1310	1320
TCAGGCTGGT	GTGATGGGAT	TCCCTGGACC	TAAAGGTGCT	GCTGGAGAGC	CCGGCAAGGC
1330	1340	1350	1360	1370	1380
TGGAGAGCGA	GGTGTTCCCG	GACCCCCTGG	CGCTGTCGGT	CCTGCTGGCA	AAGATGGAGA
1390	1400	1410	1420	1430	1440
GGCTGGAGCT	CAGGGACCCC	CTGGCCCTGC	TGGTCCCGCT	GGCGAGAGAG	GTGAACAAGG
1450	1460	1470	1480	1490	1500
CCCTGCTGGC	TCCCCCGGAT	TCCAGGGTCT	CCCTGGTCCT	GCTGGTCCTC	CAGGTGAAGC
1510	1520	1530	1540	1550	1560
AGGCAAACCT	GGTGAACAGG	GTGTTCCTGG	AGACCTTGGC	GCCCCTGGCC	CCTCTGGAGC
1570	1580	1590	1600	1610	1620
AAGAGGCGAG	AGAGGTTTCC	CTGGCGAGCG	TGGTGTGCAA	GGTCCCCCTG	GTCCTGCTGG
	1640 GCCAACGGTG				1680 CTGGTGCCCC
					1740 AACGTGGTGC
					1800 GTGCTGATGG
1810	1820	1830	1840	1850	1860
CTCTCCTGGC	AAAGATGGCG	TCCGTGGTCT	GACCGXXCCCC	ATTGGTCCTC	CTGGCCCTGC
	1880 GGTGACAAGG				
1930	1940	1950	1960	1970	1980
TCGTGGTGCC	CCCGGAGACC	GTGGTGAGCC	TGGTCCCCCC	GGCCCTGCTG	GCTTTGCTGG
1990	2000	2010	2020	2030	2040
CCCCCTGGT	GCTGACGGCC	AACCTGGTGC	Talaggegaa	CCTGGTGATG	CTGGTGCCAA
2050	2060	2070	2080	2090	2100
AGGCGATGCT	GGTCCCCCTG	GGCCTGCCGG	ACCCGTTGGA	CCCCCTGGCC	CCATTGGTAA
	CCTGGAGCCA	AAGGTGCTCG	CGGCA::CGCT	GGTCCCCCTG	GTGCTACTGG
1110001001	2180 GCTGCTGGCC	GACACGCACC	TCCTOGCCCC	TCTGGAAATG	CTGGYCCCCC
TGGCCCTCCT	2240	2250	. 2260	2270	2280
	GGTCCTGCTG	.ĢÇAAĄGAAGG	CGGCAAAGGT	CCCCGTGGTG	AGACTGGCCC
2290	2300	2310	2320	2330	2340
TGCTGGACGT	CCTGGTGAAG	TTGGTCCCCC	TGGTCCCCCT	GGCCCTGCTG	GCGAGAAAGG
2350 ATCCCCTGGT	2360 GCTGATGGTC	CTGCTGGTGC	TCCTGGTACT	CCCGGGCCTC	AAGGTATTGC
	GGTGTGGTCG	GCCTGCCTGG	TCAGAGAGGA		TCCCTOGTCT
TCCTCCCCCC	2480	2490	2500	2510	2520
	TCTGGTGAAC	CTGGCAAACA	AGGTCCCTCT	GGAGCAAGTG	GTGAACGTGG

FIG. 16B

					2600
2520	2540	2550	2560	2570	2580
2530 TCCCCCCGGT	CCCNTCCCCC	CCCTGGATT	GGCTGGACCC	CCTGGTGAAT	CTGGACGTGA
TCCCCCCGGT	CCCX1000CC		• • • • • • • • • • • • • • • • • • • •		
0.000	2600	2610	2620	2630	2640
2590 GGGGGCTCCT	2000	CTTCCCTGG	ACGAGACGGT	TCTCCTGGCG	CCAAGGGTGA
CCCCCCCC	CCICCCOOLG	0,1000.00			
	2660	2670	2680	2690	2700 GTGCCCCTGG
2650 CCGTGGTGAG	2000	CACCACCCC	TECTECTECT	GGTGCTCCTG	CICCCCICC
CCGTGGTGAG	ACCGGCCCCG	Cloncocc	10010	•	
		2210	2740	2750	2760
2710	2720	2730	ACCACCACAG	ACTIGITICATE	CTGGTCCCGC
CCCCGTTGGC	CCTGCTGGCA	ACACIOGICA	10010011000	AC.00.00.0	
		0700	2000	2810	2820
2770	2780	2790	2800	CCCCFFCCCC	CCCCTCCTCA
COGTCCCGTC	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCCCCCTGG	CCCCCCCC	CCCONGCC	CCCGTGGTGA
		2050	2060	2870	2880
2830	2840	2830	2000	CCTC CCCTC	COMPANY
CAAGGGTGAG	ACAGGCGAAC	AGGGCGACAG	AGGCAYAAAG	COLCACCOLO	GCTTCTCTGG
2890	2900	2910	2920	2730	2940
CCTCCAGGGT	CCCCCTGGCC	CICCIGGCIC	TCCTGGTGAA	CAAGGICCCT	CTGGAGCCTC
i					
2950	2960	2970	2980	2990	3000
TGGTCCTGCT	GGTCCCCGAG	GTCCCCCTGG	CTCTGCTGGT	GCTCCTGGCA	AAGATGGACT
3010	3020	3030	3040	3050	3060
CAACGGTCTC	CCTGGCCCCA	TTGGGCCCCC	TGGTCCTCGC	GGTCGCACTG	GTGATGCTGG
				•	
3070	3080	3090	3100	3110	3120
TCCTGTTGGT	CCCCCCGCCC	CTCCTGGACC	TCCTGITCCC	CCTGGTCCTC	CCAGCGCTGG
3130	3140	3150	3160	3170	3180
TTTCGACTTC	AGCTTCCTCC	CCCAGCCACC	TCAAG!.GAAG	GCTCACGATG	GTGGCCGCTA
					3240
CTACCGGGCT	agatctGCCC	TGGACACCAA	CTATICICTIC	AGCTCCACGG	AGAAGAACTG
3250	3260	3270	3280	3290	3300
CTGCGTGCGG	CAGCTGTACA	TTGACTTCCG	CAAGGACCTC	GGCTGGAAGT	GGATCCACGA
3310	3320	3330	3340	3350	3360
GCCCAAGGGC	TACCATGCCA	ACTTCTGCCT	CGGGCCCTGC	CCCTACATTT	GGAGCCTGGA
					•
3370	3380	3390	3400	3410	3420
CACGCAGTAC	AGCAAGGTCC	TGGCCCTGTA	CAACCAGCAT	AACCCGGGCG	CCTCGGCGGC
3430	3440	3450	3460	3470	3480
GCCGTGCTGC	CTCCCCCACG	CCCCCCFCCC	CCTCCCCATC	GTGTACTACG	TGGGCCGCAA
3490	3500	3510	3520	* 3530	3540
GCCCAAGGTG	GAGCAGCTGT	CCAACATGAT	CGTGCGCTCC	TGCAAGTGCA	GCTGAtctag
3550	3560	3570	3580	3590	3600
a.,.,.			4 4 4 4 4 4 5 5 5 5		2000
				· · · · · · · · · · · · · · · · · · ·	

FIG. 16C

10	20	30	40°	. 50	60 GASGPMGPRG
OLSYGYDEKS	TGCISVPCPM	CPSCPRCLPG	FRANCIQUE	Q07702.02.	120
PPGPPGKNGD	80 ΣAGKPGRP	GERGPPGPQG	ARGLPGTAGL	PCMKCHRGFS	GLEGANGING
130	140	150	160	170	180
PAGPKGEPGS	PGENGAPGQH	GPRGLPGERG	RPGAPGPAGA	RGNDGATGAA	GPPGPTGPAG
190	200	210	220		240
PPGFPGAVGA	KGEAGPQGPR	GSEGPQGVRG	EPGPPG?AGA		CQFGAKGANG
250	260	270	280	290	300
AFGLAGAPGF	PGARGPSGPQ	GPGGPPGPKG	NSGEPGAPGS	KGDTGAXGEP	GPVGVCGPPG
310	320	330	340	350	360
PAGEEGKRGA	RGEPGPTGLP	GPPGERGGPG	SRGFFGADGV	AGPKGPAGER	GSPGPAGPKG
370	380	390	400	410	420
SPGEAGRPGE	AGLPGAKGLT	GSPGSPGPDG	KTGPPGPAGQ	DGRPGPPGPP	GAP.GQAGVMG
430	440	450	460	470	460
FPGPKGAAGE	PGKAGERGVP	GPPGAVGPAG	KDGEAGAQGP	PGPAGPAGER	GEQGPAGSPG
490	500	510	520	530	540
FQGLPGPAGP	PGEAGXPGEQ	GVPGDLGAPG	PSGARGERGF	PGERGVQGPP	GPAGPRGANG
550	560	570	580	590	600
APGVDGAKGD	AGAPGAPGSQ	Gapgloghpg	ERGAAGLPGP	XGDRGDAGPX	GADGSPGKDG
610	620	630	640	650	660
VRGLTGPIGP	PGPAGAPGDK	GESGPSGPAG	PTCARGAPGD	RGEPGPPGPA	GFAGPPGADS
670	680	690	700	710	720
QPGAKGEPGD	AGAKGDAGPP	GPAGPAGPPG	PIQNV SAPGA	KGARGSAGPP	GATGFPGAAG
	740 AGPPGPPGPA				
790	008	810	820	830	840
PAGAPGTPGP	VVƏRÇƏALƏQ	GLPGQRGERG	FPGLPGPSGE	PGKQGPSGAS	GERGPPGPHG
	660 SGREGAPAAE				
910	920.	930	940	950	960
KSGDRGETGP	Agpagpvgpā	GARGPAGPQG	PRGDKGETGE	QGDRGIKGHR	GFSGLQGPPG
					1020 GDAGPVGPPG
1030 PPGPPGPPGP		1050 PQPPQEXXXD			1080 DDRDFEPSLG
	1100 LRVVQCSDLG	_		1130 KITEIKOGOF	1140 NYLANLYALI
					1200 ITKVAKVTFN
1210	1220	1230	1240	1250	1260
	GTNPLKSSGI	Engapogikak	LSYIRIADIN	ITSIPQCLPP	SLTELYLDAN

FIG. 17A

1270	1280	1290	1300	• 1310	1320	
KISRVDAASL	KGLNNLAXLG	LSFNSISAVD	NGSLANTPHL	RELHLINNKL	TRVPGGLAEH	
		,	,		•	
1330	1340	1350	1360	1370	1380	
KYIOVVYLYN						
	14125110555		,D,,504512 5	m vginbiq.	51111011110	
1300	1400	1410	1420	1430	1440	
1390	1400	1410	1420	1420	1440	
A TOLGNYK*.						

10	TGGISVPGPM	30	40	. 50	60
QLSYGYDEKS		GPSGPRGLPG	PPGAPGPQGF	QGPPGEPGEP	Gasgphgprg
70	80	90	100	. 110	120
PPGPPGKNGD	DCEAGKPGRP	GERGPPGPQG	Arglpgtagl	PGMKGHRGFS	GLDGAKGDAG
130	140	150	160	170	180
PAGPKGEPGS	PGENGAPGOM	GPRGLPGERG	RPGAPGPAGA	RGNDGATGAA	GPPGPTGPAG
190	200	210	220	230	240
PPGFPGAVGA	KGEAGPQGPR	GSEGPQGVRG	EPGPPGPAGA	Agpagnpgad	GQPGAKGANG
250	260	270	280	290	300
APGIAGAPGF	PGARGPSGPQ	GPGGPPGPKG	NSGEPGAPGS	KGDTGAKGEP	GPVGVQGPPG
310	320	330	340	350	360
PAGEEGKRGA	RGEPGPTGLP	GPPGERGGPG	SRGFPGADGV	Agpkgpager	GSPGPAGPKG
370	380	390	400	410	420
SPGEAGRPGE	AGLPGAKGLT	GSPGSPGPDG	KTGPPGPAGQ	DGRPGPPGPP	GARGQAGVMG
430	440	450	460	470	480
FPGPKGAAGE	PGKAGERGVP	GPPGAVGPAG	KDGEAGAQGP	PGPAGPAGER	GEQGPAGSPG
490	500	510	520	530	540
FQGLPGPAGP	PGEAGKPGEQ	GVPGDLGAPG	PSGARGERGF	PĠERGVQGP?	GPAGPRGANG
550	560 AGAPGAPGSQ	570	580	590	600
610 VRGLTGPIGP	620	630	640	650	660
670	680	690	700	710	720
QPGAKGEPGD	AGAKGDAGPP	GPAGPAGPPG	PIGNVGAPGA	KGARGSAGPP	GATGFPGAAG
730	740	750	760	770	780
RVGPPGPSGN	AGPPGPPGPA	GKEGGKGPRG	ETGPAGRPGE	VGPPGPPGPA	GEKGSPGADG
790 PAGAPGTPGP	800	810	820	830	840
850 PPGLAGPPGE	860	870	880	890	900
	920	930	940	950	960
	980	990	1000	1010	1020
	1040	1050	1060	1070	1080
1090		1110	1120	1130	1140

FIG. 18

9 18 27 26 45 50 CAG CIG TOT TAT GGC TAT GAT GAG ANA TCA ACC CGA GGA ATT TCC GTG CCT GGC CCC ATG GOT CCC TOT GOT CCT CGT GOT CTC CCT GGC CCC CCT GGT GCA CCT GGT CCC CAA GGC TTC 139 147 ' 156 CHA GGT CCC CCT GGT GAG CCT GGC GAG CCT GGA GCT TCA GGT CCC ATG GGT CCC CGA GGT CCC CCA GGT CCC CCT GGA ANG ANT GGA GAT GAT GGG GAA GCT GGA ANA CCT GGT CGT CCT GGT GAG, CGT GGG CCT CCT GXG CCT CAG GGT GCT CCA GGA TTG CCC GGA ACA GCT GGC CTC CCT GGA ANG MAG GGA CAC AGA GGT TTC AGT GGT TTG GAT GGT GCC AAG GGA GAT GCT GGT CCT GCT GGT CCT ANG GCT GAG CCT CGC AGC CCT GGT GAA AAT CGA GCT CCT GGT CAG ATG 429 438 447 456 465 474

GGC CCC CGT GGC CTG CCT GGT GAG AGA GGT CGC CCT GGA GCC CCT GGC CCT GGT GCT GCT COT GGA AAT GAT GOT ACT GGT GCT CCC GGG CCC CCT GGT CCC ACC GGC CCC GCT GGT CCT CCT GGC TIC CCT GGT GCT GGT GGT GCT AAG GGT GAA GCT GGT CCC CAA GGG CCC CGA GGC TOT GAN GGT CCC CAG GGT GTG CGT GGT GAG CCT GGC CCC CCT GGC CCT GCT GGT GCT CCT CCC CCT CCT CCA AAC CCT CCT CCT GAT CGA CAG CCT CCT GCT AAA CCT CCC AAT CCT GCT CCT GCT ATT GCT CGT GCT CCT GGC TIC CCT GGT GCC CGA GGC CCC TCT GGA CCC CAG UGC CCC GGC GGC CCT CCT GGT CCC AAG GGT AAC AGC GGT GAA CCT GGT GCT CCT CGC AGC AAA OGA GAC ACT GGT CCT AAG OGA GAG CCT GGC CCT GTT GGT GTT CAA GGA CCC CCT GGC CCT GCT GGA GAG CAA GGA AAG CGA GGA GCT CGA GGT GAA CCC GGA CCC ACT GGC CTG CCC SCA CCC CCT GGC GAG CGT GGT CGA CCT GGT AGC CGT GGT TTC CCT GGC GCA GAT GGT GTT GCT GGT CCC AAG CGT CCC GCT GGT GAA CGT GGT TCT CCT GGC CCC GCT GGC CCC AAA GGA TOT COT OUT GAA GOT GOT COT COT GAA GOT GOT CTG COT GGT GCC AAG GGT CTG ACT COA AGO OCT GGC AGO CUT GGT CCT GAT GGC AAA ACT GGC CCC CCT GGT CCC GGT CAA 

FIG. 19A

1.105 1269 1278 1267 1296 THE COTT GOAL OUT ANA COTT GOT GOT GOT GOS GAG COOLOGS AND GOT GOAL GAG COALGOT GOT GOT GGA CCC COT CUC GCT GTC GGT CCT GCT GCC AAA GAT GGA GAG GCT GGA GCT CAG GGA CCC CCT GGC CCT GGT GGT CCC GCT GGC GAG AGA GGT GAA CAA GGC CCT GCT GGC TCC CCC GGA TTC CAG GGT CTC CCT GGT CCT GGT GGT CCT CCA GGT GAA GCA GGC AAA CCT GGT GAA CAG GGT GTT CCT GGA GAC CTT GGC GCC CCT GGC CCC TCT GGA GCA AGA GGC GAG AGA GGT TTC CCT GGC GAC CGT GGT GTG CAA GGT CCC CCT GGT CCT GGA CCC CGA GGG GCC AAC GGT GCT CCC GCC AAC GAT GCT GCT AAG CGT GAT GCT GGT GCC CCT GGA GCT CCC GGT AGC CAG GGC GCC CCT GGC CTT CAG GGA ATG CCT GGT GAA CGT GGT GCA GCT GGT CTT CCA GGG CCT .1758 ARG GGT GAC AGA GGT GAT GCT CGT CCC AAA GGT CCT GAT GGC TCT CCT GGC AAA GAT GGC 1845 1854 ONE CON COT CITE ACC GRE CCC ATT GOT CCT CCT GGC CCT GCT GCC CCT GGT GAC AAG GGT GAA AGT GGT CCC AGC GGC CCT GGT GGT CCC ACT GGA GCT CGT GGT GCC CCC GGA GAC CAA CCT CCT GCT AAA GGC GAA CCT GGT GAT CCT GGT GCC AAA GGC GAT GCT GGT CCC CCT COG CCT GCC GGA CCC GCT GGA CCC CCT GGC CCC ATT GGT AAT GTT GGT GCT CCT GGA GCC ANA OUT GOT COC COC AGO COT GOT CCC COT GOT GCT ACT GOT TTC CCT GGT GCT GCT GGC CGA GTC GGT CCT CGT CGC CCC TCT GGA AAT GCT GGA CCC CCT GGC CCT CCT GGT CCT GCT 2238 · 2247 GGC AAA GAA GGC CGC AAA GGT CCC CGT GGT GAG ACT GGC CCT GCT GGA CGT CCT GGT GAA GTT GGT CCC OUT GGT CCC CCT GGC CCT GGT GGC GAG AAA CGA TCC CCT GGT GCT GAT GGT CCT OCT GOT CCT CCT GOT ACT CCC COG CCT CAA GOT ATT GCT GGA CAG CGT GOT GTG GTC GGC CTC CCT CAG AGA GGA GAG AGA GGC TTC CCT GGT CTT CCT GGC CCC TCT GGT GAA COT COC ARA CAR GOT CCC TOT GGA GCA AGT GGT GAR CGT GGT CCC CCC GGT CCC ATG GGC 

FIG. 19B

2529 2538 2547 2556 2.05 2574 CCC CCT GGA 17G CCT GGA CCC CCT GGP GAA TCT GGA CGT GAG GGG GCT CCT GCT GCC GAA GGT TCC CCT GGA CAC GGT TCT CCT GGC GCC AAG GGT GAC CGT GAG ACC GGC CCC AND ACT COT CAT COT CAT CAR ACT COT COT COT GOT CCC GCT CCC GCT CCC GCC CCC 2757 2796 GCC GCC CGT GCC CCC GCC GGA CCC CAA GGC CCC CGT GGT GAC AAG GGT GAG ACA GGC GAA CAG GGC GAC AGA GGC ATA AAG GGT CAC CGT GGC TTC TCT GGC CTC CAG GGT CCC CCT GGC CUT COT GGC TOT COT GGA CAA CGT CCC TOT GGA GCC TOT GGT CCT GGT CCC CGA GGT CCC CCT GGC TCT QCT GGT GCT CCT GGC AAA GAT GGA CTC AAC GGT CTC CCT GGC CCC ATT GGG CCC CCT GGT CCT CGC GGT CGC ACT GGT GAT GCT GGT CCT GTT GGT CCC CCC GGC CCT CCT GGA CCT CCT GGT CCC CCT GGT CCT CCC AGC GCT GGT TTC GAC TTC AGC TTC CTC CCC CAG CCA CCT CAA GAG AAG GCT CAC GAT GGT GGC CGC TAC TAC CGG GCT AGA TCC GAT GAG GOT TOT GGG ATA GCC COA GAA GIT COT GAT GAC CGC GAC TTC GAG CCC TCC CTA GGC CCA GIN; TGC CCC TITC CGC TGT CAA TGC CAT CTT CGA GTG GTC CAG TGT TCT GAT TTG GGT CTG GAC AAA GTG CCA AAG GAT CTT CCC CCT GAC ACA ACT CTG CTA GAC CTG CAA AAC AAC ALA ATA ACC GAA ATC AAA GAT GGA GAC TTT AAG AAC CTG AAG AAC CTT CAC GCA TTG ATT CTT GTC AAC AAT AAA ATT AGC AAA GIT AGT CCT GGA GCA TTT ACA CCT TIG GTG AAG TTG GAA CGA CTT TAT CTG TCC AAG ANT CAG CTG AAG GAA TTG CCA GAA AAA ATG CCC AAA ACT CIT CAG GAG CTG COT GCC CAT GAG AAT GAG ATC ACC AAA GTG CGA AAA GTT ACT TTC AAT CGA CTG AAC CAG ATG ATT CTC ATA GAA CTG CGC ACC AAT CCG CTG AAG AGC TCA GGA ATT GAN ANY GOO OCT THE CAS GON AND AND CHE TEE THE ATE COE ATT GET GAT ACE ANT ATC ACC AGC ATT CCT CAA GGF CIT CCT CCT TCC CTT ACG GAA TTA CAT CTT GAT GGC AAC

FIG. 19C

223	378	9	375	§ T. CC2	3907	CTY:	3916	ርፕር	3825 TAA TAA	TIG (	3834 GCT AAG	TTG GGA
ויייי			911 05		oc. noc	C15	Act Och		7411 7511			
	384	9	385	8	3867		3876		3885		3894	
TTG	AGT TT	O.K. 0	AGC AT	C Wi	CCT GTI	GAC	AAT CCC	TCT	CLC CCC	AAC	ACG CCT	CAT CTG
									3945			
AGG	GÄG CT	ን ፒ ርአር	TTG GA	e C AAC	AAC AAG	CTT	ACC AGA	GTA	CCT GGT	GGG (	CTG GCA	GAG CAT
	396	9	397	8	3997		3996		4005		4014	
AAC	TAC AT	C CAG	CITY CI	u tac	CTT CAT	AAC	ARC RAT	ATC	TCT GTA	GTI (	GGA TCA	AGT GAC
	402	0	403	ů	4047		4056		4065		4074	
"ተጥር"	402 402	ን እ	GC 4 C	ር ይልረ	APA TOS	AAG	GCT TCT	TAT	TCG GGT	GTG	AGT CTT	TIC AGC
110	102 00		00.5 07	- FE-10-	70.0 1001							
	408	9	409	3	4107		4116		4125		4134	
AAC	CCG GT	ב כאם	TAC TO	₹ GAG	ATA CAG	CCA	TCC ACC	TTC	AGA TGT	GTC '	TAC OTG	CGC TCT
		^	43.6		4167		4176		4185		4194	
000	414	y • ~~~	91:	ιή 	4107		4110		4100		4134	
GCC	ATT CA	A CTC	GGA A	C. TAI	WAY JAY	• • •	•••		• • • • • • • • • • • • • • • • • • • •		• • • • • •	• • • • • •

FIG. 19D

. 10 ggaaggatt	20 tccatttccC	30. AGCTGTCTTA	40 TGGCTATGAT	50 GAGAAATCAA	06 TAADDADDOO
	08 DOTACCOCO	90	100	110	120
	140 CAAGGCTTCC		1.00	120	180
			770	230	240
TCCCATGGGT	CCCCGAGGTC	CCCAGGICC	200	290	300
TGGAAAACCT	GGTCGTCCIG	CICHCCOICC	000.00.000		
GCCCGGAACA	320 GCTGGCCTCC	CIGORITON	oodicitoria.		
	380 GATGCTGGTC	300	400	410	420
		450	460	470	480 GCCCTGGAGC
400	500	510	520	530	540
CCCTGGCCCT	CTGGTGCTC	GTGGAAATGA 570	580	590	GGCCCCCTGG 600
TCCCACCGGC	CCCGCTGGTC	CICCIGGCII	CCCTGGTGCT	GIIGGIGCIA	NOOCH TOWNS
1CGTCCCCAA	620 GGGCCCGAG	630 GCTCTGAAGG	640 TCCCCAGGGT	650 GTGCGTGGTG	AGCCTGGCCC
670	680 CCTGGTGCTG	690 CTGGCCCTGC	700 TGGAAACCCT	710 GGTGCTGATG	720 GACAGCCTGG
					780 CTGGTGCCCG
790	800	810	820	830	ACÁGCGGIGA 840
850	860	870	880	890	900
ACCTGGTGCT	CCTGGCAGCA	AAGGAGACAC	TGGTGCTAAG	GGAGAGCCTG	GCCCTGTTGG
TGTTCAAGGA		CTGCTGGAGA	GGAAGGAAAG	CGAGGAGCTC	GAGGTGAACC
970 CGGACCCACT	· 980 GGCCTGCCCG	990 GACCCCCTGG	1000 CGAGCGTGGT	1010 GGACCTGGTA	1020 GCCGTGGTTT
1030 CCCTGGCGCA	1040 GATGGTGTTG	1050 CTGGTCCCAA	1060 GGGTCCCGCT	1070 GGTGAACGTG	1080 GTTCTCCTGG
	1100 CCCAAAGGAT	1110 CTCCTGGTGA	1120 AGCTGGTCGT	1130 CCCGGTGAAG	1140 CTGGTCTGCC
1150 TGGTGCCAAG	1160 CGTCTGACTG	1170 GAAGCCCTGG	1180 CAGCCCTGGT	1190 CCTGATGGCA	1200 AAACTCGCCC
1210 CCCTCGTCCC	1220 CCCGGTCAAG	1230 ATGGTCGCCC	1240 CGGACCCCCA	1250 GGCCCACCTG	1260 GTGCCCGTGG

FIG. 20A

	•		-		
1270 TCAGGCTGGT	1280 GTGATGGGAT	1290 TCCCTGGACC	1300 TAAAGGTGCT	. 1310 GCTGGAGAGC	1320 CCGGCAAGGC
1330 TGGAGAGCGA	1340 GGTGTTCCCG	1350 GACCCCCTGG	1360 CGCTGTCGGT	1370 CCTGCTGGCA	1380 ADADOTADAA
1390	1400 CAGGGACCCC	1410	1420	1430	1440
1450 CCCTGCTGGC	TCCCCCGGAT	1470 TCCAGGGTCT	CCCTGGTCCT	CCTGGTCCTC	1500 CAGGTGAAGC
1510 AGGCAAACCT	1520 GGTGAACAGG	1530 GTGTTCCTGG	1540 AGACCTTGGC	1550 GCCCCTGGCC	1560 CCTCTGGAGC
1570 AAGAGGCGAG	1580 AGAGGTTTCC	1590 CTGGCGAGCG	1600 TGGTGTGCAA	1610 GGTCCCCCTG	1620 GTCCTGCTGG
1630 ACCCCGAGGG	1640 GCCAACGGTG	1650 CTCCCGGCAA	1660 CGATGGTGCT	1670 AAGGGTGATG	1680 CTGGTGCCCC
1690	1700 GGTAGCCAGG	1710	1720	1730	1740
1750	1760 CCAGGGCCTA	1770	1780	1790	1800
1810	1820 Aaagatggcg	. 1830	1840	1850	1860
	1880				
TOGTGCCCCT	GGTGACAAGG	GTGAAAGTGG	TCCCAGCGGC	CCTCCTCGTC	CCACTGGAGC
1930 TCGTGGTGCC	1940 CCCGGAGACC	1950 GTGGTGAGCC	1960 TGGTCCCCCC	1970 GGCCCTGCTG	1980 GCTTTGCTGG
1990 CCCCCCTGGT	2000 GCTGACGGCC	2010 AACCTGGTGC	2020 TAAAGGCGAA	2030 CCTGGTGATG	2040 CTGGTGCCAA
2050	2060 GGTCCCCCTG	2070	2080	2000	3100
2110	2120	2130	2140	2150	21.60
	CCTGGAGCCA				
2170 TTTCCCTGGT	2180 GCTGCTGGCC	2190 GAGTCGGTCC	TCCTGGCCCC	· 2210 TCTGGAAATG	2220 CTGGACCCCC
2230 TGGCCCTCCT	2240 GGTCÇTGCTG	2250 GCAAAGAAGG	2260 CGGCAAAGGT	2270 CCCCGTGGTG	2280 AGACTOGCCC
2290 TCCTGGACGT	2300 CCTGGTGAAG	2310 TTGGTCCCCC	2320 TGGTCCCCCT	2330 GGCCCTGCTG	2340 GCGAGAAAGG
2350 ATCCCCTGGT	2360 GCTGATGGTC	2370 CTGCTGGTGC	2:380 TCCTGGTACT	2390	2400 AAGGTATTGC
2410		2430	2440	2450	2460
2470	2480	2490	2500	2510	

FIG. 20B

					•	
253	0	2540	2550	256	0 . 2570	2580
TCCCCCCC	T	CCCATGGGC	CCCCTGGAT	r GGCTGGACC	C CCTGGTGAAT	CIGGACGIGA
259	90	2600	2610	262	0 2630	2640
GGGGGCTCC	T	GCTGCCGAAG	GTTCCCCTG	ACGAGACGG	T TCTCCTGGCG	CCAAGGGTGA
269	50	2660	2670	268	0 2690	2700
CCCTCCTC	٩G	ACCGGCCCCC	CTGGACCCCC	TGGTGCTCN	T GGTGCTCNTG	GTGCCCCTGG
27:	10	2720	2736	274	0 2750	2760
CCCCGTTCC	30	CCTGCTGGC	A AGAGTGGTG	A TCGTGGTGA	g acregrecie	CTGGTCCCGC
27	70	2780	2790	280	0 2810	2820
CGGTCCCG1	rc	GCCCCCCCT	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	cccccccc	A CCCCAAGGCC	CCCGTGGTGA
283	30	2840	2850	286	0 2870	2880
CAAGGGTGJ	٩G	<b>ACAGGCGAAG</b>	: AGGGCGACAG	3 AGGCATAAA	G GGTCACCGTG	GCTTCTCTGG
289	90	2900	2910	292	0 2930	2940
CCTCCAGG	T	CCCCCTGGCC	: CICCIGGCIA	C TCCTGGTGA	A CAAGGTCCCT	CTGGAGCCTC
299	50	2960	2970	2986	0 2990	3000
recrete	T	GGTCCCCGAC	GTCCCCCTGC	CICIGCIGG	r GCTCCTGGCA	AAGATGGACT
301	LO	3020	3030	3040	3050	3060
CAACGGTCT	rc	CCTGGCCCC	TTGGGCCCCC	TOGTCCTCGC	GGTCGCACTG	GTGATGCTGG
307	0	3080	3090	3100	3110	
TCCTGTTGG	T	CCCCCCGGCC	CTCCTGGACC	TCCIGGICCO	CCTGGTCCTC	3120
212	^	20.0				ccyccc 100
TTTCGACTT	~	3140	3150	3160	3170	3180
-	_		CCCACCACC	TURAGAGAAG	GCTCACGATG	GTGGCCGCTA
319	0	3200	3210	3220	3230	2240
- TACCGGGC	T	agatetecaa	AGGATCTTCC	CCCTGACACA	3230 ACTCTGCTAG	ACCTGCAAAA
325	n	3260	2020			
AACAAAAT;	Α.	ACCGAAATCA	AAGATGGAGA	CTTTAAGAAC	3290 CTGAAGAACC	3300
721/		334-			annance	LICACGCATT
ATICTIGICA ATICTICICA	, : 1	021L ««««TACAA	3330	3340	3350	3360 ·
	•		LINGCAAAGT	TAGTCCTGGA	3350 TAActgcag.	* * * * * * * * * * *

FIG. 20C

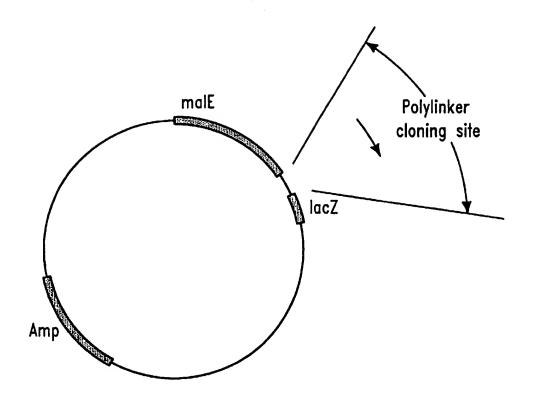


FIG. 21

MALE ATC GAG GGA AGG ATT TCA GAA TTC GGA TCC TCT AGA GTC GAC CTG CAG GCA AGC TTG LACZ			
:			
Ě		 	
AGC		Hindli	
CCA	•		
CAG		<b>-</b> -	
CTG		₽ + •	
SAC	'	_	
) DI		- B S	
GA G	.	_	
CT A	1	D O	Č
20	1	•	CC 714
3A T	BamHi		<u> </u>
) )	m	·	_
A	EcoRl		
A GA	БО		
75			
ATT	С		
AGG	E		
GGA	×		
GAG			
ATC			
	Sp. ·		
MAL			

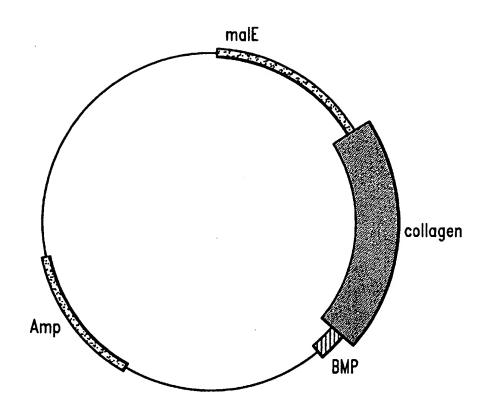


FIG. 23

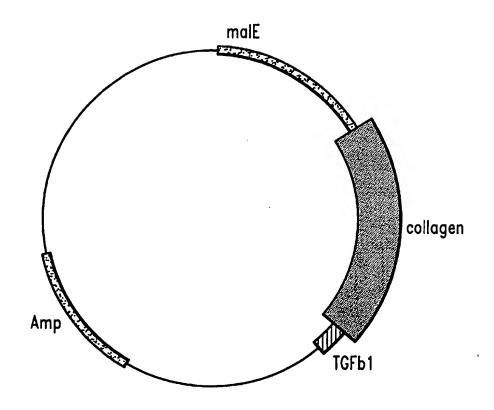


FIG. 24

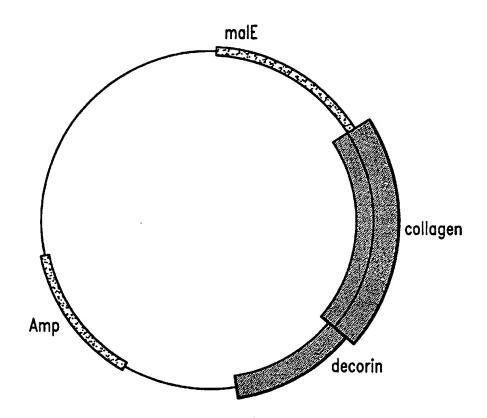


FIG. 25

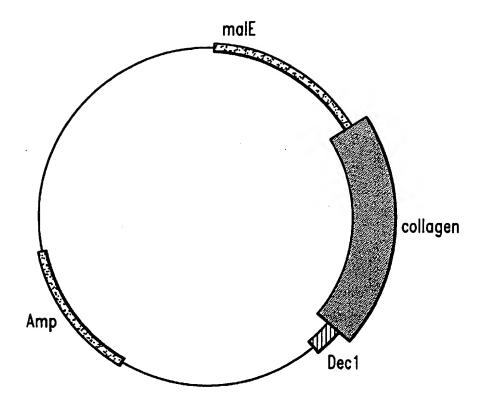


FIG. 26

	ಘ್ರ		9	æ:-	~~~	18	ር <b>ኔ</b> ቸ	GAG.	27 AAA	TCA	ACC	36 GGA	GGA	ATT	45 700	GTG	cct	5.4 GGC
5.	<u>ت</u> بر	CTG	lCi	In.	GLU	 Tu~	Asn	Glu	Lvs	Ser	 Thr	Gly	Gly	īle	Se:	Val	Pro	Gly
											•	90			99			108
	ccc	ATG	CGT	<u>cc</u>	TCT	GGT	CCT	CGT	<u> </u>	CTC	CCI	GGC	ccc	<u> </u>	GGT 	GCA	CCT	GGT
	320	met	GŢĀ	9:0	Ser	GŢÀ	Pro	Arg	Gly	Leu	Pro			5:0			210	Giy 162
	ccc	CAX	117 GGC	TTC	CAA	126 GGT	ccc	CCT	135 GG7	G),G	CCT	144 GGC	G7G	œ:	153	GCT	7C3	GGT
																		Gly
			,171		~~``	180	· ~~	CCZ	189	CY.	CCT.	198	AAG	AAT	207 GG3		GAT	216 GGG
			_									~						
	?=0	Met			Arg			Pro	243		2.0	252			251			Gly 270
	ججي	. GCT	225	eaa .	. cc:	234 GGT	Œ	CCI	G	GHG	ಡಾ			CC:			CX	G
	Glu	Ala	Gly	Lys	520	Gly	Arg	Pro	Gly	Glu	بد	Gly	Pro	820	Gly	?::0	Glr	s GŢĀ
			279	رميد.	٠,	288	303	ب	297		: 007	306 GG3		ARG	315		AC:	324 . GGT
												·						Gly
	~c	وخم	333			342			351			360		-30	369		•	378
	770	AGT			GAT	-		AAG			GCT			œ			230	G
	Phe	Ser	GLy	Leu	çe.	Gly	Ala	Lys	ŒĴå	çzá	Ala	Gly	Pro	Ala	Gly	Pro	Ŀŷs	Cly
	GAG	CCT	387 GGC	22A	CCT	396 GGT		. aat	405 GGA		cci	414 GGT		ats	423 GGC		CGT	432 GGC
	Glu	 Pro	Glv	Ser	220	Glv	Glu	 Asa	Glv	Ala	 Fro	Giv	Gla	Met	G17		aro	Gly
			441			450			459			468			477		•	456
	CTG	CCT	GGT	GAG	AGA	GGT	<u> </u>	cci	CEA	$\frac{1}{2}$	CCI	GGC	CCT	GCT	GGT.	GCT	œ	CCY.
	Leu	Pro	Gly	Glu	λrç	Gly	Arg	?ro	Gly	Ala	Fro	Gly	920	Ala	Giy	<u>Ala</u>	ÿzş	Gly
	AAT	GAT	495 GGT		ACT	504 GGT	GCT	GCC	513 666	œ	CCT	522 GGT	œ	эС.	531	~~	ينئ	540 GGT
																		Gly
			549		••-	558			567		•••	576		•••	585			594
	CCT	CCT		TTC	ಯ	GGT		GIT	ŒĨ			GGT			Œ		CXX	GGG
	920	520	Gly	Phe	Pro													
	ccc	CGA	603 GGC	TCT	GAA	612 GGT	ccc	CAG	621 GGT	GTG.	CGT	630 GGT	GAG	~~	639	רנר	CCT	648 GGC
		Arg																
		7	657			666	-:-		675		. <b>-</b> y	624			62y 693		0	702
	CC:	GCT		GC:	GCT		CCI	GCT		AAC	CCT		GCT	GAT		CAG	CCT	
	Pro	ALa	Gly	λla	Ala	GŢÀ.	210	Ala	Gly	Asa	?::	GŢĀ	Nε	ès:	<u></u>	51:	?:0	GTA.

FIG. 27A

711 720 729 738 747 75 GCT ANA GGT GCC ANT GGT GCT GCT GGT ATT GCT GGT GCT GCC TTC CCT GG
Ala Lys Gly Ala Asm Gly Ala Pro Gly Ele Ala Gly Ala Pro Gly Phe Pro Gly
765 774 783 792 801 81 GCC CGA GGC CCC TCT GGA CCC CAG GGC CCC GGC GGC CCT CCT GGT CCC AAG GG
Ala Arg Gly Pro Ser Gly Pro Gin Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly
819 828 837 846 855 86
AND AGO GGT GAA CCT GGT GCT CCT GGC AGO AAA GGA GAC ACT GGT GCT AAG GGE Asn Ser Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly
120 AAD AAD AAD TOO TOO TOO DOO TOO AAD TTO TOO TOO TOO DOO TOO DAA
Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu Glu Gly 927 936 945 954 963 972
AAG CGA GGA GCT CGA GGT GAA CCC GGA CCC ACT GGC CTG CCC GGA CCC CCT GGC
Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro Gly Pro Pro Gly
981 990 999 1008 1017 1026 GAG CGT GGT GGA CCT GGT AGC CGT GGT TTC CCT GGC GGA GAT GGT GTT GCT GGT
Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly Ala Asp Gly Val Ala Gly
1035 1044 1053 1062 1071 1080 CCC ANG GGT CCC GGT GGT GGA CGT GGT GGT CCT GGT CCC GGT GGC CCC GGT GGC CCC ANA GGA
Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser Pro Gly Pro Ala Gly Pro Lys Gly
1089 1098 1107 1116 1125 1134 TOT COT GGT GAA GGT GGT GGT GGT GGT GGT GGT GG
Ser Pro Gly Giu Ala Gly Arg Pro Gly Giu Ala Gly Leu Pro Gly Ala Lys Gly
1143 1152 1161 1170 1179 1188 CTG ACT GGA AGC CCT GGC AGC CCT GGT CCT GAT GGC AAA ACT GGC CCC CCT GGT
Leu Thr Gly Ser Pro Gly Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly
1197 1206 1215 1226 1222 1222
are
Pro Ala Gly Gln Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly 1251 1260 1269 1278 1287 1296
CAG GCT GGT GTG ATG GGA TTC CCT GGA CCT ANA GGT GCT GCT GGA GAG CCC GGC
Gln Ala Gly Val Mer Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly
AAG GCT GGA GAG CGA GGT GTT CCC GGA CCC CCT GGC GCT GTC GGT CCT GGC
Lys Ale Giy Glu Are Gly Val Pro Gly Pro Pro Gly Ale Val Gly Pro Ale Gly

FIG. 27B

				٠,	262		1	177		1	1386		:	1395			1404
		אננו	CAG	رحت ا	200	GCT	CAG	Œλ	$\alpha$	CCT	GGC	CCI	CCT	333	CCC	GCT	GGC
*/*/*	نذذ	المتان															
ive	250	GLV	Glu	Ala	Gly	Ala	Gln	Gly	5:0	510	Gly	520	λla	GŢĀ	520	λιa	Giv
273	وت.	:			•									1449			1458
		1413		1	L422			1431	<b>T</b>	m.	1440	TTC	CAG	GGT	CIC	CCT	GGT
GAG	λGλ	CGT	СУY	Cya	GGC	cci	GCT		1W					GGT			
						950	Ala	Gly	Ser	210	Glv	Phe	Gla	Gly	Leu	220	CŢÀ
Glu	Arg	GŗĀ	GIu	GIN	GTÅ	220											
		1 467		:	1476			1485			1494			1503			1512
C	GCT	GGT	CCT	CCA	GGT	CYY	GCA	Œ	AAA	cci	GGT	CAA	CXG	GGT	GTT	CUI	المثان
											~~~	Clu	Gla	Gly	Va!	2=0	Giv
?:0	λLa	Gly	520	5:0	CŢĀ	Glu	Aia	GIĀ	TÀR	120	GTĀ	بادق	<b>G1.</b> .	Q1,			Giy
					1530			1539			1548			1557			1566
		1521	٠	CCT	GGC	ccc	TOT	CEY.	GCA	AGA.	GGC	ಆರ	AGA	GG	777	CCI	GGC
Aso	Leu	Gly	Ala	Pro	Gly	Pro	Ser	. CīĀ	Ala	Arg	GLY	Glu	yzg	G1Ã	Phe	310	Gly
		•												1611			1620
		1575			1584			T237	~	-	1007	~	CC2	-02	GCC		GGT
G;G	CGT	G	GiG	CAA	(60)												
Clu	2	· Glu	Val	Gin	Glv	Pro	Pro	Glv	210	Ala	Gly	Pro	Arg	Gly	Ala	Asn	GŢĀ
Gio		, 019	•••								•						
		1629	•		1638			1647			1656			1665			1674
GCT	. cc:	: ccc	AAC	GAT	. GGI	. ec:	AAG	G	GYI	GCT	GGT	GCC	œ	CCA	GC:	سنت	GGT
							7		2==	112	Gly	212	2-0	Glu	داد	2-0	Gly
عند	720	Giy	ASR	وعم	Grà	A.c	. rys	Gry	بردم		GIY	ALE		CLY			U-3
		1683			1692	!		1701			1710						1728
AGO	CAC	GG	: σα	CCT	GGC	CTI	CAG	GGA	ATG	CCT	GGT	CAA	CGT	GGT	GCA	GCT	CCT
		. <u></u> .															
Ses	Gli	. Gly	YTS	Pro	Gly	Leu	Gla	Gly	140	Pro	GŢĀ	Glu	Arg	Gly	λļē	علد	Gly
								1762			1764			1773			1782
	CC3	1737	~	320	1746	620	962	1755 ???	CZT	GT	1764 GGT	ccc		GGI			
Leu	250	Gly	220	Lys	Gly	Asp	وعد	GŢ.	ديد	<u> 212</u>	Gly	220	Lys	Gly	<u> 212</u>	Asp	Gly
				•		-											
		1791			1800						1818			1827			1836
707	27.	CCC	AAA	GAI	GGC	GIC	CGT	G-	C.G	تبلغ	GGC	ατ	ATT	GGT	α:	CUT	حت
Sar	2-0	Gl-	fue	300	Gly	Va1	3	Gly	Ten	The	Gly	D-0	Tia	Gly	2	2-0	63.4
Je.		زين	ny 3	بردء	GIJ	*	~~	uny	~~-	***	GTĀ	0		GIŽ		220	227
		1845			1854						1872			1881			1890
cci	GCT	GGT	$\alpha$	CCT	GGT	GAC	aag	CGT	G33	agt	GGI	$\alpha$	AGC	$\alpha$	CCT	CCT	GT:
520	A!a	GŢĀ	yya	Pro	GľÃ	ysb	Lys	GIA	Giu	Sez	<b>GT</b> Ā	510	Ser	G17	523	ALA	GŢĀ
		1899			1908		,	1617			1925		,	1935		,	944
-0-									620			GEG		GGI	~		
														Gly	Pro	920	Gly
		•		•	•			•	•	_	_			·			•
		1953			1962			1971			1980	<b>.</b>		1989			.998
CCI	CCT	GGC	TTT	CCT	GGC	CCC	CCT	GT	GCI	G;C	GGC	CAA	CI	GGT	CCT	AA.	GGC
0	215	Glo	Dho	A1-	G1	D	D	Glu	113	Δε	Glu	Gla	P=-	Gly	A1=	145	Gly
3	¢	GTÅ	: ue	ي در	GTÅ	-10	-10	A	مند	اردم	Gry	CLII	0	GTĀ	عند	ră2	o±y.
		2007			2016		:	2025			2034		2	2043		2	052
GAA			GAT									CCC		GGG	CC:		
, <b></b> -																	
Glu	3.20	Gly	çZń	Alz	Gly	Ala	Lys	GLY	Αsp	ala.	GŁy	2:5	320	Gly	?=0	عند	Cī 'n

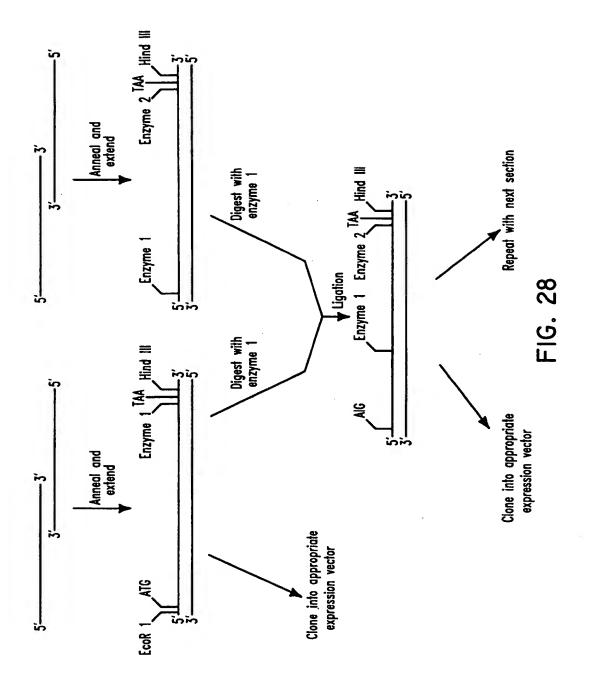
FIG. 27C

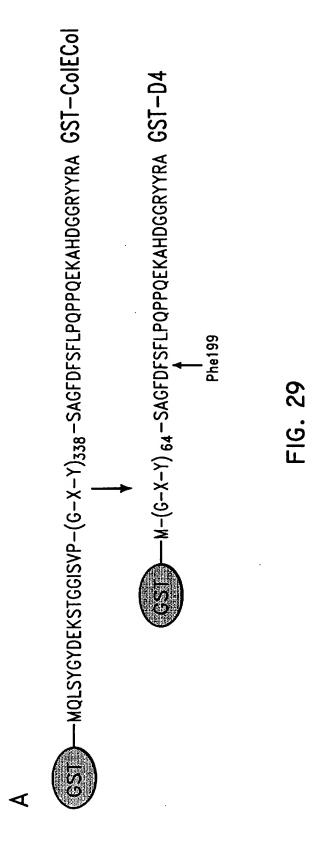
2061	2070		207	}	20	880		:	2097			2106
CCC CCI GGY CCC	CCT GGC	CCC .	ATT GG	TAA	GTT (	CGT	GCT	CCT	GGA	GCC	AAA	CGT
Pro Ala Gly Pro	?ro Gly	510	Ile Gl	/ Asn	Val (	Gly	Ala	Pro	CTĀ	ALA	rāz	GIĀ
•									2151			2160
2115	2124	000	213	به درست ۱	3CT /	142	<b>444</b> C	CT.	CC1	GCT		
OCT OGC GGC AGC	GCT GGT	CCC	CCT GG	r GC1	النام	Col	110	<u></u>				
Ala Arg Gly Ser	22- 22-		0 c C)	- 212	The	Glu	Phe	Pro	Glv	Ala	λla	Gly
Ala Arg Gly Ser	YIS CIÀ	P.20	210 01	, ,,,,,	****	013			2			•
2169	2178		<b>218</b>	7	2	196			2205			2214
<i>CEY CLC CCL CCL</i>	CCT GGC	CCC	TCT GG	A AAT	GCT	GGA	CCC	CCT	GGC	CCT	CCI	GGT
Arg Val Gly Pro	Pro Gly	Pro	Ser Gl	y Asn	Ma	CJĀ	510	Pro	Gly	510	Pro	GīĀ
2223	2232		224	<u> </u>	2	250	~~~		2259			2268
CCT GCT GGC AAA	GRA GGO	GGC	AAA G			GGI	C-10	MC:				
Pro Ala Gly Lys	Clu Clu	Glu	tus G	v Pro	3-0	Glv	Glu	The	Glv	Pro	Ala	Glv
PEO ALE GLY LYS	GIU GLY	Gry		,		G-y	<u></u>		٠			1
2277	2286		229	5		304			2313			2322
CGT CCT GGT G24	GTT GGT	$\infty$	cc: cc	: CCC	CCT	೦೦೦	CCT	CCT	GCC	GłG	AAA	GGA
Arg Pro Gly Glu	Val Gly	Pro	2:0 G1	A Bro	Sto	Gly	Pro	Ala	Gly	Glu	Lys	GŢĀ
2331	2240		234	۵.	•	358			2367			2376
2331 TCC CCT GGT GCT							ACT					
100 001 001												
Ser Pro Gly Ala	Aso Glv	220	Ala Gl	v Ala	220	Glv	Thr	Pra	Glv	320	Gla	G1y
2385	2394		240	3	2	412		٠.	2421			2430
2385 ATT GCT GGA CAG	2394		240	3	2	412		٠.	2421			2430
ATT GCT GGA CAG	2394 CGT GGT	GTG	240 GTC GG	3 C CIG	CCT	412 GGT	CAG	AGA	2421 GGA	GAG	AGA	2430 GGC
2385 ATT GCT GGA CAG Ile Ale Gly Gin	2394 CGT GGT	GTG	240 GTC GG	3 C CIG	CCT	412 GGT	CAG	AGA	2421 GGA	GAG	AGA	2430 GGC
ATT GCT GGA CAG Ile Ala Gly Gin 2439	2394 CGT GGT Arg Gly 2448	GTG Val	240 GTC GC Val Gl	3 C CTG  y Leu 7	CCT C	412 GGT  Gly 466	CAG GLn	AGA Arg	2421 GGA Gly 2475	GAG Glu	AGA Arg	2430 GGC Gly 2484
ATT CCT CGA CAG	2394 CGT GGT Arg Gly 2448	GTG Val	240 GTC GC Val Gl	3 C CTG  y Leu 7	CCT C	412 GGT  Gly 466	CAG GLn	AGA Arg	2421 GGA Gly 2475	GAG Glu	AGA Arg	2430 GGC Gly 2484
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC GCT GGT CTT	2394 CGT GGT Arg Gly 2448 CCT GGC	GTG Val	240 GTC GG Val Gl 245 TCT GG	G CTG  Leu  GAA	CCT (	GLy 466 GGC	CAG Gln AAA	AGA Arg CAA	2421 GGA Gly 2475 GGT	Glu CCC	AGA Arg	2430 GGC Gly Gly 2484 GGA
ATT GCT GGA CAG Ile Ala Gly Gin 2439	2394 CGT GGT Arg Gly 2448 CCT GGC	GTG Val	240 GTC GG Val Gl 245 TCT GG	G CTG  Leu  GAA	CCT (	GLy 466 GGC	CAG Gln AAA	AGA Arg CAA	2421 GGA Gly 2475 GGT	Glu CCC	AGA Arg	2430 GGC Gly Gly 2484 GGA
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly	GIG Val CCC	Val Gl TCT GS TCT GS	G CTG  CTG  Y Leu  G CAA	2 CCT (	G1y 466 GGC G1y G1y	CAG Gln AAA	ACA Arg CAA Gln	GLy GGT GLy GLy	Glu Glu CCC	ACA Arg TCT Ser	2430 GGC Gly 2484 GGA Gly
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502	GIG Val CCC	240 GTC GG Val Gl 245 TCT GG Ser Gl	3 C CIG Y Leu 7 T GAA Glu	2 CCT ( CCT ( Pro ( 2)	412 GGT Gly 466 GGC Gly	CAG Gln AAA Lys	ACA Arg CAA Gln	2421 GGA Gly 2475 GGT Gly 2529	GAG Glu CCC	ACA Arg TCT Ser	2430 GGC Gly 2484 GGA Gly 2538
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502	GIG Val CCC	240 GTC GG Val Gl 245 TCT GG Ser Gl	3 C CIG Y Leu 7 T GAA Glu	2 CCT ( CCT ( Pro ( 2)	412 GGT Gly 466 GGC Gly	CAG Gln AAA Lys	ACA Arg CAA Gln	2421 GGA Gly 2475 GGT Gly 2529	GAG Glu CCC	ACA Arg TCT Ser	2430 GGC Gly 2484 GGA Gly 2538
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT	GTG Val	240 GTC GG Val Gl 245 TCT GG Ser Gl 251 CCC GG	GAA	CCT (CCT (Pro (CCT (CCT (CCT (CCT (CCT (CCT (CCT (CC	412 GGT Gly 466 GGC Gly 520 GGC	CAG Gln AAA Lys	AGA Arg CAA Gln CCT	2421 GGA Gly 2475 GGT Gly 2529 GGA	GAG Glu CCC Pro	ACA Arg TCT Ser	2430 GGC Gly 2484 GGA Gly 2538 GGA
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly	GTG Val	240 GTC GG Val GI 245 TCT GG Ser Gl 251 CCC GG	GAA	CCT (CT (CT (CT (CT (CT (CT (CT (CT (CT	412 GGT Gly 466 GGC Gly 520 GGC	CAG Gln AAA Lys	ACA Arg CAA Gln CCT Pro	2421 GGA Gly 2475 GGT Gly 2529 GGA GGA Gly	GAG Glu CCC Pro	ACA Arg TCT Ser GCT	2430 660 61y 2484 66A 61y 2538 66A 61y
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556	CCC Pro	240 GTC GG Val Gl 245 TCT GG Ser Gl 251 CCC GG Pro Gl	GA GLU Pro	CCT (CT (CT (CT (CT (CT (CT (CT (CT (CT	412 GGT Gly 466 GGC GIY 520 GGC GIY	CAG Gln AAA Lys CCC	AGA Arg CAA Gln CCT Pro	2421 GGA Gly 2475 GGI Gly 2529 GGA GIY 2583	CCC Pro	ACA Arg TCT Ser GCT	2430 GGC Gly 2484 GGA Gly 2538 GGA GLY 2538
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556	CCC Pro	240 GTC GG Val Gl 245 TCT GG Ser Gl 251 CCC GG Pro Gl	GA GLU Pro	CCT (CT (CT (CT (CT (CT (CT (CT (CT (CT	412 GGT Gly 466 GGC GIY 520 GGC GIY	CAG Gln AAA Lys CCC	AGA Arg CAA Gln CCT Pro	2421 GGA Gly 2475 GGI Gly 2529 GGA GIY 2583	CCC Pro	ACA Arg TCT Ser GCT	2430 GGC Gly 2484 GGA Gly 2538 GGA GLY 2538
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA	CCC Pro	240 GTC GG Val Gl 245 TCT GG Ser Gl 251 CCC GG Pro Gl 256 GAG GG	GAA  GAA  GEAA  GE	CCT (CT (CT (CT (CT (CT (CT (CT (CT (CT	412 GGT 466 GGC GGL GGL GGC GGC GGC	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro	2421 GGA Gly 2475 GGI GGY 2529 GGA GGY 2583 GGT	GAG Glu CCC Pro TTG Leu	AGA Arg TCT Ser GCT Ala	2430 GLy 2484 GGA GLy 2538 GGA GLy 2592 GGA
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA	CCC Pro	240 GTC GG Val Gl 245 TCT GG Ser Gl 251 CCC GG Pro Gl 256 GAG GG	GAA  GAA  GEAA  GE	CCT (CT (CT (CT (CT (CT (CT (CT (CT (CT	412 GGT 466 GGC GIV 520 GGC GGC GGC	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro	2421 GGA Gly 2475 GGI GGY 2529 GGA GGY 2583 GGT	GAG Glu CCC Pro TTG Leu	AGA Arg TCT Ser GCT Ala	2430 GLy 2484 GGA GLy 2538 GGA GLy 2592 GGA
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601	2394 CGT GGT Arg Gly 2446 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610	CCC Pro	Z40 GTC GG Val Gl Val Gl Z45 TCT GG Ser Gl Z51 CCC GG Pro Gl GAG GG GAG GG	GOA CTG	2 CCT ( Pro / Pro / Pro /	412 GGT Gly 466 GGC GGC GGC GGC GGC GGC GGC GGC GGC G	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro GAA Glu	2421 GGA Gly 2475 GGI GGI 2529 GGA GGI GGI 2583 GGI GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GG	CCC Glu CCC Pro TTG Leu TCC	AGA Arg TCT Ser GCT Ala CCT	2430 632 61y 2484 632 61y 2538 632 61y 2592 634 61y
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu	2394 CGT GGT Arg Gly 2446 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610	CCC Pro	Z40 GTC GG Val Gl Val Gl Z45 TCT GG Ser Gl Z51 CCC GG Pro Gl GAG GG GAG GG	GOA CTG	2 CCT ( Pro / Pro / Pro /	412 GGT Gly 466 GGC GGC GGC GGC GGC GGC GGC GGC GGC G	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro GAA Glu	2421 GGA Gly 2475 GGI GGI 2529 GGA GGI GGI 2583 GGI GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GGI 4863 GG	CCC Glu CCC Pro TTG Leu	AGA Arg TCT Ser GCT Ala CCT	2430 632 61y 2484 632 61y 2538 632 61y 2592 634 61y
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601  CGA GAC GGT TCT	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610 CCT GGC	GTG Val Val CCC Pro CCC Arg	240 GTC GG Val Gl Val Gl 245 TCT GG Ser Gl 251 CCC GG Pro Gl 256 GAG GG	GEACA	2 CCT (CT (CT (CT (CT (CT (CT (CT (CT (CT	412 GGT GGLy 466 GGC GGC GGC GGC GGC GGC GGC GGC GGC G	CAG Gln AAA Lys CCC Pro	AGA Arg Gln CCT Pro GAA GLu	2421 GGA 	CCC Pro TTG Leu TCC Ser	AGA Arg TCT Ser GCT Ala CCT Pro	2430 61y 61y 2484 61y 61y 2538 61y 2592 664 61y 2664 6664
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610 CCT GGC	GTG Val Val CCC Pro CCC Arg	240 GTC GG Val Gl Val Gl 245 TCT GG Ser Gl 251 CCC GG Pro Gl 256 GAG GG	GEACA	2 CCT (CT (CT (CT (CT (CT (CT (CT (CT (CT	412 GGT GGLy 466 GGC GGC GGC GGC GGC GGC GGC GGC GGC G	CAG Gln AAA Lys CCC Pro	AGA Arg Gln CCT Pro GAA GLu	2421 GGA 	CCC Pro TTG Leu TCC Ser	AGA Arg TCT Ser GCT Ala CCT Pro	2430 61y 61y 2484 61y 61y 2538 61y 2592 664 61y 2664 6664
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601  CGA GAC GGT TCT  Arg Asp Gly Ser	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610 CCT GGC	GTG Val Val CCC Pro CCC Arg	240 GTC GG Val Gl  Val Gl  Z45 TCT GG Ser Gl  251 CCC GG GAG GG GAG GG GAG GG GAG GG Lys Gl  Lys Gl	GAC ASP	2 CCT ( Pro	Gly 466 GGC Gly 520 GGly 574 GGIY 574 GGIY GGIY GGIY GGIY GGIY GGIY GGIY GGI	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro GAA GLu ACC	2421 GGA GLy 2475 GGI 2529 GGA GLY 2533 GGT GLY 2533 GGT GLY 2533 GGT GLY 2533 GGT GLY 2533 GGT GLY 2544 GGT GGY 2554 GGT GGY 2554 GGY GGY GGY GGY GGY GGY CGY GGY G	CCC Pro TTG Leu TCC Ser	AGA Arg TCT Ser GCT Ala 2 CCT Ala Ala	2430 GCC GLY 2484 GCA GLY 2538 GCA GLY 2592 GGA GLY 2646 GCA GLY 2646 GCA GCA GCA GCA GCA GCA GCA GCA
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601  CGA GAC GGT TCT  Arg Asp Gly Ser  2655	2394 CGT GGT Arg Gly 2446 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610 CCT GGC	GTG Val CCC Pro CCC Pro CCT Arg	Z40 GTC GG Val Gl Val Gl Z45 TCT GG Ser Gl Z51 CCC GG Z56 GAG GG GAG GG GAG GG GAG GG Lys Gl	GAC  GAC  GAC  Asp	2 CCT ( Pro A Met ( CCT ) Arg ( CCT )	412 GGT GLy 466 GGC GGC GGLy GGG GGC GGG GGG GGG GGG GGG GGG GGG GG	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro GAA Glu ACC	2421 GGA GIY 2475 GGIY 2529 GGA GIY 2529 GGA GGIY 637 GGIY 637 GGIY 637 637 637 637 637 637 637 637	CAG Glu CCC Pro TTG Leu TCC Ser	AGA Arg TCT Ser GCT Ala 2 CCT Pro GCT Ala	2430 6Gly 22484 6Gly 2538 6Gly 2538 6Gly 2592 6Gly 2646 6Gly 2646 6Gly 2700
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601  CGA GAC GGT TCT  Arg Asp Gly Ser	2394 CGT GGT Arg Gly 2446 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610 CCT GGC	GTG Val CCC Pro CCC Pro CCT Arg	Z40 GTC GG Val Gl Val Gl Z45 TCT GG Ser Gl Z51 CCC GG Z56 GAG GG GAG GG GAG GG GAG GG Lys Gl	GAC  GAC  GAC  Asp	2 CCT ( Pro A Met ( CCT ) Arg ( CCT )	412 GGT GLy 466 GGC GGC GGLy GGG GGC GGG GGG GGG GGG GGG GGG GGG GG	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro GAA Glu ACC	2421 GGA GIY 2475 GGIY 2529 GGA GIY 2529 GGA GGIY 637 GGIY 637 GGIY 637 637 637 637 637 637 637 637	CAG Glu CCC Pro TTG Leu TCC Ser	AGA Arg TCT Ser GCT Ala 2 CCT Pro GCT Ala	2430 6Gly 22484 6Gly 2538 6Gly 2538 6Gly 2592 6Gly 2646 6Gly 2646 6Gly 2700
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601  CGA GAC GGT TCT  Arg Asp Gly Ser  2655	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610 CCT GGC Pro Gly	GCC GCC Alla GCCT GCCC GCCC GCCC GCCC GCCC GCCC GCC	240 GTC GG Val Gl  245 TCT GG Ser Gl  251 CCC GG Pro Gl  256 GAG GG GAG GG  Lys Gl  267 Lys Gl  267 CCT GG	GAC Asp	2 CCT ( Pro A Pro A Arg ( CCT ( C))))))))))	412 GGT G1y 466 GGC G1y 520 GGC GGC G1y 574 GCT G1y GGC GGC GGC GGC GGC GGC GGC GGC GGC GG	CAG Gln AAA Lys CCC Pro GCC Ala GAG GAG	AGAA Gln CCT Pro GAA Glu ACC Thr	24211 GGA GLy 24755 GLy 25299 GGA GLy 25833 GGT GLy 25833 GGT GLy 25833 GGT GLY 25836 GLY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836 GGY 25836	CAG Glu CCC Pro TTG Leu TCC Ser CCC	AGA Arg TCT Ser GCT Ala CCT Pro Ala GCT Ala	2430 61y 61y 22484 63x 61y 2538 61y 2538 61y 2592 66x 61y 26x 61y 2592 66x 61y 2592 66x 61y 2592 66x 67x 67x 67x 67x 67x 67x 67x

FIG. 27D

. 2709			754
aag agt gat gat	t cot cot cag act cot	CCT CCT CCC CCC CGT CCC GTC	GGC
Lys Ser Gly As	o Arg Gly Glu Thr Gly	Pro Ala Gly Pro Ala Gly Pro Val	Gly
2763	2772 2731	2790 2799 2	808
CCC SCT GGC GCX	ASS 300 300 300 TOO 3	CCC CAA GGC CCC CGT GGT GAC AAG	GG:
Pro Ala Gly Ala	a Arg Gly Pro Ala Gly	Pro Gla Gly Pro Arg Gly Asp Lys	CŢĀ
2817	2825 2835	2844 2853 2	862
CAG ACA GGC GAS	A CAG GGC GAC AGA GGC	ATA AAG GGT CAC CGT GGC TTC TCT	GGC
Glu The Gly Glu	e Gla Gly Asp Arg Gly	Ile Lys Gly His Arg Gly Phe Ser	GJÀ
2871	2880 2889	2898 2907 2	916
CTC CAS GGT CCC	CET GGC CCT CCT GGC	TOT COT GGT GAA CAA GGT CCC TOT	CGj
Leu Gin Gly Pro	s to Gly Pro Pro Gly	Ser Pro Gly Glu Gln Gly Pro Ser	Gly
2925	2934 2943	2952 2961 2	970
		CCC CCT GGC TCT GCT GGT GCT CCT	SSC
Ala Ser Gly Pro	o Ala Gly Pro Arg Gly	Pro Pro Gly Ser Ala Gly Ala Pro	2ī'y
2979	2988 2997	3006 3015 3	024
AAA GAT GGA CTO	C AAC GGT CTC CCT GGC	CCC ATT GGG CCC CCT GGT CCT CGC	GGT
Lys Asp Gly Le	i Ash Gly Leu Pro Gly	Pro Ile Gly Pro Pro Gly Pro Arg	Gly
3033	3042 3051	3060 3069 3	078
		े राठा राठा स्टब्स राठा राठा ठवन राठा राठा	GGT
Arg The Gly Ass	Ala Gly Pro Val Gly	Pro Pro Gly Pro Pro Gly Pro Pro	<u> 31</u> y
3087	3096 3105	3114 3123 3	132
		CAC TTC AGC TTC CTC CCC CAG CCA (	
Pro Pro Gly Pro	Fro Ser Ala Gly Phe	Asp Phe Ser Phe Leu Pro Gln Pro	?ro
2141	3150 3159	3168	
	क्य क्या क्या क्या क्या		
Gim Glu Lys Ala	His Asp Gly Gly Arg	Tyr Tyr Arg Ala	

FIG. 27E





	HCol	ColECol	_
Proline			
CCU	139	11	
CCC	93	12	
CCA	6	27	
CCG	0	189	
Glycine			
GGU	174	147	
GGC	97	179	
GGA	64	8	
GGG	11	12	

FIG. 30

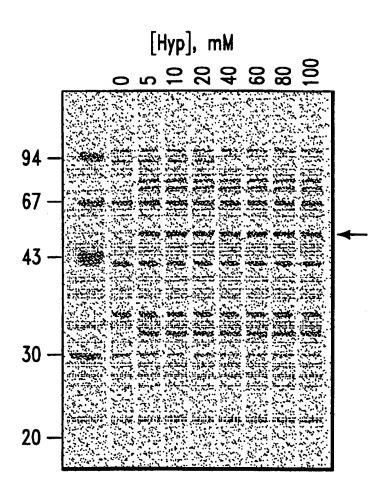


FIG. 31

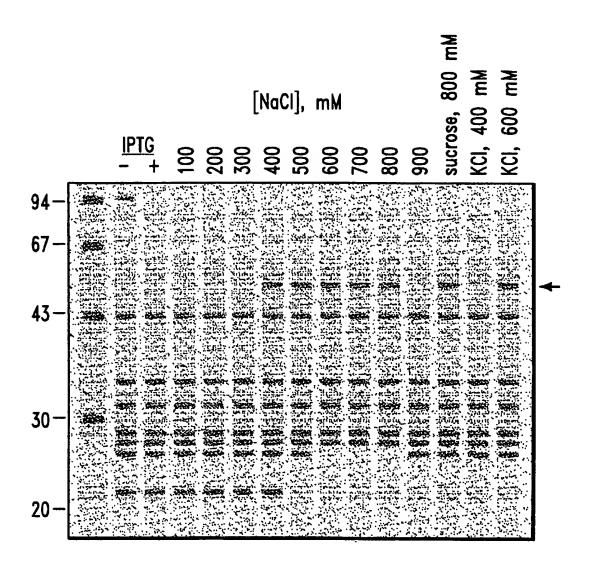
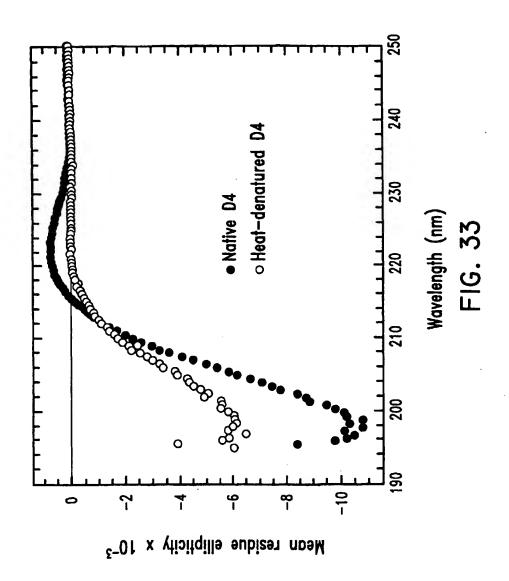


FIG. 32



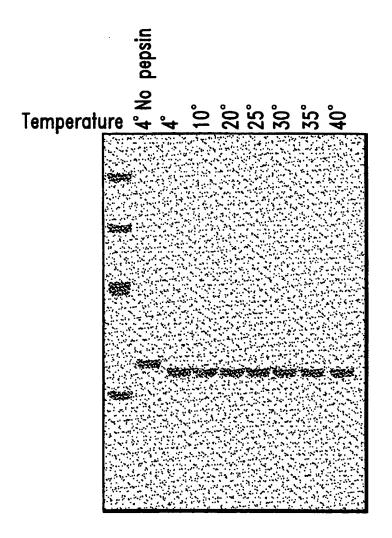


FIG. 34

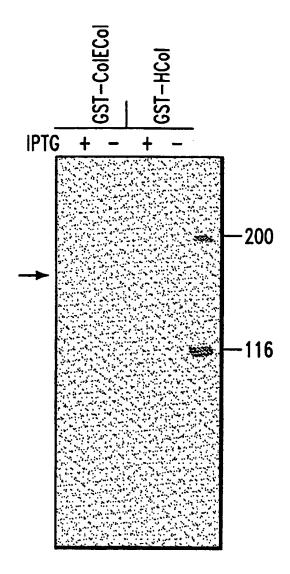


FIG. 35

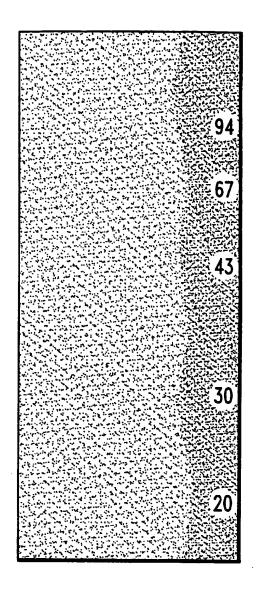


FIG. 36

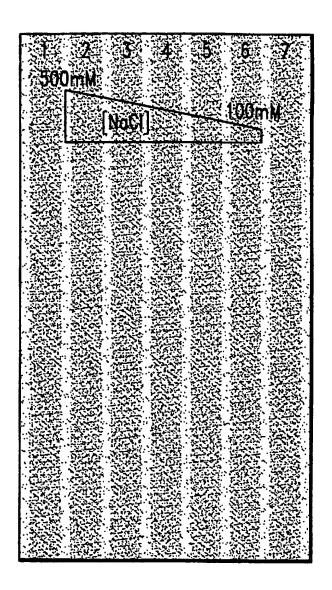


FIG. 37

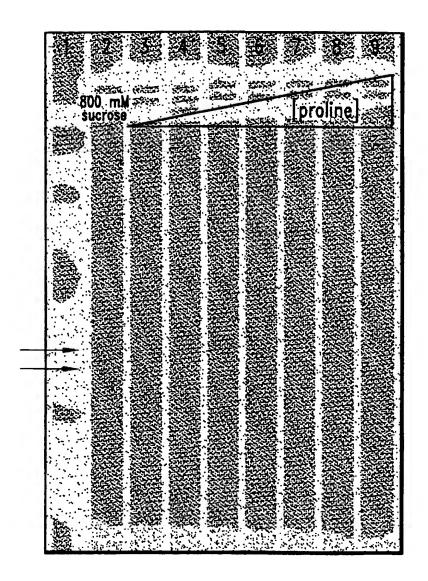


FIG. 38

5'	CAG	CTG	. 9 .xcc	ΤΑΤ	GGC	18 TAT	GAT	GAA	27 AAA	AGC	ACC	36 660	GGC	AIC	45 AGC		CCG	54 GGC
,																		
	Gln	Leu	Ser	Tyr	GIY	Tyr	YSO	GIU	rys	Ser	Inr	GI.	GIĀ	116	ser	val	PIO	Gly
	~~~	N.T.C	63	œ	ACC.	72	œ	CGT	81	CTG	ccs	90 GGC	CCG	CCA.	99 GGT		σcc	108 GGT
	Pro	Mec	Gly	Pro	Ser	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Pro	Pro	GŢĀ	Ala	Pro	Gly
			117			126			135			144			153			162
	CCG	CAG	GGC	TTT	CAG	GGT	<u></u>	ccc	GGC	GAA	ccc	GGC	GAA	CT	GGT	ecce	AGC	GGC
	Pro	Gln	Gly	9r.e	Gln	Gly	Pro	Pro	Gly	Glu	Pro	Gly	Glu	Pro	Gly	Ala	Ser	CJ À
			171			180			189			193			207			216
	CCG	ATG	GGC	œ	CGC	GGC	CCG	$\infty$	GGT	$\infty$	CCA	GCC	AAA	A-iC	GGC	GAT	GAT	GGC
	Pro	Mec	Gly	Pro	Arg	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Lys	Asa	Gly	Asp	Asp	Gly
			225		·	234			243			252			261			270
	GAA	GCG	GGC	AAA	$\infty$	GGA	CGT	CCG			CGT		$\infty$	$\infty$	GGC	$\infty$	CAG	CCC
	Glu	Ala	Gly	Lys	Pro	Gly	Arg	Pro	GJA	Glu	Arg	Gly	Pro	Pro	Gly	Pro	Gln	Gly
•			279			288			297			305			315			324
	ထေ	CGC	GGA	ĊĭG	$\infty$			ငေ			ccc		atg	AAA		CAC	œc	
	Ala	Arg	Gly	Leu	Pro	Gly	Thr	Ala	GŢĀ	Leu	Pro	GŢĀ	Met	Lys	Gly	His	Arg	Gly
			333			342			351			360			369			378
	TTC	TCT		CTG	CAT		GCC	AAA		GAC	CCC		CCG	ಯ	GGT	ငင္သ	aaa	
	Phe	Ser	Glv	Leu	Asp	Gly	Ala	Lys	Gly	λsp	λla	Gly	Pro	Ala	Gly	Pro	Lys	Gly
			387		_	396		•	405	_		414			423			432
	GAG	$\alpha$		AGC	CCG			AAC		ထေ	CCC		CAG	atg		CCG	ŒT	
	Glu	Pro	Glv	Ser	Pro	Gly	Glu	Asn	Gly	Ala	Pro	Gly	Gln	ᇩ	Gly	Pro	Arg	Gly
			441			450			459			463			477			486
	CTG	CCT		GAA	œ		CGC	<b>∞</b>		GCC	$\infty$		CCA	CT.	-	GCA	ŒĨ	
							3				 D		 D			11-		
	Leu	PIQ	_	GIU	Arg		Arg	510		ALG	P20		FIO	nia		Ala	Atg	
	244	CZT	495	œ	204	504 GGT	ന്ദ	ന	513	42)	m	522 CCC	ന്നു	275	531	CCG	æ	540 CCT
	Asn	Asp	CJÀ	Ala	Thr	Gly	Ala	Ala	Gly	Pro	Pro	Сĵå	Pro	Thr	Gly	Pro	Ala	Gly
			549			558			567			575			585			594
	ccc	ccc	œc	TTT	<u> </u>	CCI.	<u></u>	eze.	GGT 	<u> </u>	AAA	GGC	GAA	CCA	GGT	222	CAG	GGG
	Pro	Pro	Gly	Phe	Pro	Gly	Ala								Gly	Pro	Gln	Gly
			603			612			621			630			639			648
	CCG	CGC	GGG	AGC	CAG	CGT	CCT	CYC	CGC	GTT	CCT	GGT	GAA	<u></u>	GGC	CCG	<b>σσ</b>	GGC
	Pro	Arg	Gly	Ser	Glu	Gly	Pro	Gln	Gly	Val	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly
			657			666			675			684			693			702
	CCG	CCC		CCC	GCG		CCC	ÇCT		AAC	CCT		ccc	CAC		CAG	CCA.	
	Pro		610	A)a	Al a	G) v	Pro	Ala	Glv	Asn	Pro	Glv	Ala	Eso.	Glu	 Gln	Pro	G) u
			;			:			-,		- 20	!			;	~		Y

FIG. 39A

											738			747			756
CXCG .	<b>///</b> /	711 GGT	CCC	MC	720 GGC	ငင္သ	CCG	CCT	ati	GCA	GG:	CCA	نت	ಆಬ	::TC	CCG	GGT
Ala			 ala	asn	Giv	Ala	510	G; Y	Ile	Ala	Gly	Ala	220	Gly	Phe	Pro	Gly
								202			702			801			810
GCC	ccc	765 GGC	ငင	:CC	774 GGC	ccs	CYC	ಹ	œ	GGC	œ	ÇCG	ccc	GGC	CCG	ልጹሕ	GGG
			 Pr3	Ser	27.A	Pro	Gln	G; y	Pro	Gly	Gly	Pro	750	Gly	320	Lys	Gly
								0:7			RES			855			864
እኣC	AGC	619 GGT	CAL	CCG	eze cgt	GCG	CCG	650	AGC	AAA	GGC	CAC	ACC	CGT	GCG	AAA	GGT
															: Ala		
					222			EGI			900			909			918
GAA	ccc	873 GGC	CCA	GTG	GGT	GTT	CAA	. œ	ccc	CCG	GGC	CCG	GCG	GGC	GAG	G53	GGC
Glu	Pro	Gly	Pro	Val	Gly	Val	Gln	Gly	Pro	?20	Gly	Pro	Ala	Gly	Glu	Glu	Gly
		927			936			945			954			963			972
AAA	œ	eg:	CC	CGC	द्धा	GAA	. 000	(CC)	COG	300	GGC	CIG	C:	GGC	CCG	CCG	GGA
Lys	Arg	Gly	Ala	Arg	Gly	Glu	Pro	G <u>i</u> y	Pro	Thr	Gly	Leu	?:0	Gly	520	Pro	Gly
		981			990	1		999	)		1003			1017			1026
GAA	CGT	· 657	œ	coc	<b>331</b>	AGO	<u>ccc</u>	GGT	777	000	SSS	GCG	GAT	GGI	GIG	GCG	GCC
Glu	Arg	GŢĀ	G1;	220	Sly	Sec	Arg	Gly	Phe	Pro	G7Å	,Ala	وعد	Gly	Val	Ala	Gly
		1035			1044			1053			1062			1071			1080
CCG	AAA	GET	<u>cc</u>	CCG	SGI	GAA	CGT		ACC	CCG	GGC	CCG	SCG	GGC	CCA	AAA 	GGC
520	Lys	Gly	Pro	Alz	Gly	Glu	Arg	GIA	Ser	2ro	Gly	Pro	Ala	Gly	?:0	Lys	Gly
		1089			1098			1107			1116			1125			1134
AGC	CCG		G	GCA	GGA	CG:	ccc	GT.	GRA	GCG	GG:	CTC	cc	GGC	CCC	AAA	GGT
Ser	Pro	GJ A	GŢ::	Ala	Gly	Yzā	Pro	сīÀ	Glu	YJS	GŢĀ	Leu	?50	Gly	λla	Lys	Gly
		1143			1152			1161			1170			1179			1188
			`	÷					~~~						<b>200</b> G		
Leu	Thr	GĵÀ	Ser	Pro	Gly	Ser	510	GŢĀ	Pro	Asp	GŢĀ	Lys	Chr	GŢĀ	5z'o	510	Gly .
~~		1197	C10		206	~~		1215	~~		224	~~		233	~~		242
															<del></del>		
Sro	Ala	Gly	GLA	Asp	Gly	Arg	Pro	Gly	520	Pro	Gīy	Pro	320	Gly	Ala	Arg	Gly
CZC		1251 GGT		ATT										287	GAA		296
G±n	ALS	GTÀ	٧٤٤		_	Phe	Pro	GTÀ	510	Lys	GŁY	YTS	ķΙā	GŢĀ	Glu	510	GIA
AAA		1305 GC	G22		1314 GGT	GTC		GGT	ജ		.332 GGC	CCT		341 GGG	CCG		.350 GGC
											<del></del> -						
Lys		•	GTA	•	_	۰۵۰		_	FZO		-	wra		•	Pro		
AAA		1359 SSC	C:A		1368 GGC	ಆಯ		1377 GGC	ထာ		.356 GGA	CCA		395 GGT	CCG		.404 GGC
															Pro		
-13	- 25	- y		, L. C	~-3	4 min (ii)	·	-y		0	G-y	240	ميده	GAŞ	220	WT.	GTÀ

FIG. 39B

								•			440		,	249		,	458
	CCC	413		225	422	ccc	CC.F	.431 .GGC	YCC.	CCS	GGT	770	CAG	GGT	CTG		
GAG	CCC	GGT	(:AH												 [	250	Glv
Glu	Arg	Gly	Glu	Gln	GΪΆ	6L0	Ala	ΟŢΆ	Ser	5.53	GIÅ	2.te	GIN	GIA	Dec	113	Ory.
	,	457		)	476		1	1485		!	494			1503		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	1512
CCT	CCG	GGT	CCA	ÇCG	GGT	GAA	GCG	CCC	222	ccs	<u> </u>	GiA	CAA	GG1	910		
	Ala	G) v	Pro	Pro	Gly	Glu	Ala	GTÀ	Lys	Pro	Gly	GĽ:u	Gln	Gly	Val	220	CŢÀ
								1 6 7 0			: 548			1557			1566
	CTG	1521	ccc	CCA	1530 GGC	ccc	AGC	03C	ccc	CGC	GGC	C÷A	œc	GGT	TTC	CCG	GGC
GAC											Glu	Gin	Arm	Glv	Phe	Pro	Glv
Asp	Leu	Gly	Ala	Pro	GIY	āzo	Ser	GLY	WT.	Mig	GLY	G_0					Gly
		1575			1584		~~~	1593	~~		1602	coc	CCC	1611 GGC	GCC		1620 GGC
	CGT																
Glu	. Arg	Gly	Val	Gln	Gly	Pro	Pro	Gly	?ro	Ala	Gly	320	Arg	Gly	Ala	Asn	GJA
		1629	١		1638	}		1647			1656			1665			1674
GCC	; ccc	GCC	AAC	CAT	GG1	CCC	AAA	GG:	GAT	ccc	GGT	GCC	CCA	GGT	GCC	CCG	GGC
A1:			ASD	Aso	Gly	Ala	Lys	Gly	Asp	Ala	Gly	عند	Pro	Gly	Ala	Pro	Gly
A.I.														1719			1725
»CC	CAG	1683 GGC	; exx	: 000	1694	: ; CIC	CAR	, egg	ATG	CCG	1710 GGT	CAA					
Se:	- Glr	GL	/ ALa	Pro	GIZ	Leu	اللكا ا	: СТА	æc	220	Grā	, Œu	Ary	GTĀ	ALG	ALG	Gly
		1737			1746			1755			1764			1773			1782
CTA	CCG	GGT	CO	. <u></u>	GGC			GGT	(F41)	ست							
Lev	. Pro	Gly	Pro	Lys	Gly	ysp	Arg	Gly	Asp	Als	Gly	?:0	Lys	Gly	Ala	Asp	Gly
		1791			1800	)		1809			1918			1827			1836
TCC	: CCT			CAT	CCC	GII	CCE	CCT	CTG				ATC	GCC	$\infty$	$\infty$	GGC
Se	Pro	Glv	Lvs	Aso	Gly	Val	Ard	Glv	Leu	The	Gly	220	Ile	Gly	Pro	Pro	Gly
		_	•	•							-			•			-
ccc	GCA	1845 . GGT			1854 GGT			1853 GGT			1872 GGT	ccs		1881 GGC	CCA	-	1890 GGC
Pro	Ala	Ciy	Ala	Pro	GLY	' ASO	Ly's	CTÅ				=10	Ser	GIŞ	Pro	Ala	GIA
		1899			1908			1917						1935		-	1944
	ACT	GGT	GGG	<b>G</b> T	GGT				GAC	<b>G</b> .	GGT	GRA	<del></del>	GGT			GGC
Pro	Thr	Gly	Ala	Arg	Gly	Ala	Pro	Gly	Asp	Arg	Gly	Glu	Pro	Gl'n	Pro	Pro	Сĵ
		1953		;	1962		. :	1971		1	1980		1	1989		1	1998
ccc	GCG																CCC
Pro	Ala	Gly	Phe	Ala	Gly	510	Pro	Gly	Ala	Aso	Glv	Gin	Pro	Gly	Ala	Lvs	Glv
		_						_		•						•	_
GAA	ccc	2007 GGG			2016 GGT			2025 GGC	GAC		GGT	ထ		2043 GGC	CCI		9052 GGC
GIU	Pro	GTÅ	ASD	ATS	GTÅ	ATS	r'ns	пñ	wż	ALE	GTÅ	220	LL0	GIÀ	FIO	ALA	GLY
		2061			2070			2079	330		8809			2097	~~~		106
	ccc						A.T		M4C	643	ان 				·	AAA 	GGT
Pro	Ala	Gly	Pro	Pro	Gly	Pro	lle	Gly	Asa	نے۷	Gly	λla	Pro	Gly	λla	Lys	GΣΥ

FIG. 39C

				2151 2160
2115	2124	. CCC 655 656 2133	2142 ACC GGT TTO	CCC GGT GCG GGG
Ala Arg Gly Ser	Yī9 CJĀ 520	Pro Gly Ala	The GLY she	ero Giy Ala Ala Giy
2169	2178	2187	2196	2205 2214
				229 233 233 235 235
Arm Val Gly Pro	Pro Gly Pro	Ser Gly Asn	Ala Gly Pro	Pro Gly Pro Pro Gly
0003	2232	2241	2250	2259 2258
FAA COD DOD DOD	නෙ ගෙ ගෙ	CCC TEO AAA C	CGT CGT CAA	ACC SEC CCT GCG GGA
pro Ela Gly Lys	Glu Gly Gly	Lys Gly Pro	Arg Gly Glu	The Gly Pro Ala Gly
			2304	2
2277 CGT CCA GGT GAA	GIG GGT CCC	ccs 665 ccs	CCC GGC CCC	GCG GGC GAA AAA GGT
				Ala Gly Glu Lys Gly
Ard his gra gra			•	
2331 ACC CCG GGT GCG	2340 GAT GGT CCC	2349 : GCC GGT SCG	2358 CCA GGC ACG	2367 2376 COG GGT CCG CAA GGT
-		ALE GIY ALE	PES GLY 1.12	Pro Gly Pro Gln Gly
2395	2394	2403		2421 2430 CSC SGC GBA CGC GGC
Ile Ala Giy Glr.	Arg Gly Val	l Val Gly Leu	Pro Gly Gin	Arg Gly Glu Arg Gly
2439	2449	2457	2466	2475 2484
TTT CCG GGT C:G		. war (g): gap		CAG GGT OCA TOT GGC
Phe Pro Gly Leu	Pro Gly Pro	Ser Gly Glu	Pro Gly Lys	Gin Gly Pro Ser Gly
2493	2502	2511	2520	2529 2538
GCG AGC GGT GAA	CGT GGC CCG	003 661 600	ATG GGC CCG	CCC GGT CTG GCG GGC
Ala Ser Gly Glu	Arg Gly Pro	Pro Gly Pro	Met Gly Pro	Pro Gly Leu Ala Gly
2547	2556	2565	2574	2583 2592
CCT CCG GGT GAA	AGC CGT CGT	GFY CCC CCC	೦೦೦ ಆರಾ ಅ೦೦	GAA GGC AGC CCA GGC
Pro Pro Gly Glu	Ser Gly Arg	Glu Gly Ala	Pro Gly Ala	Glu Gly Ser Pro Gly
2601	2610	2619	2628	2637 2646
				NCC 635C 005C 635C
Arg Asp Gly Ser	Pro Gly Ala	Lys Gly Asp	Arg Gly Glu	Thr Gly Pro Ala Gly
2655		2673		•
		CCC GCT GCC	2682 CCA GGC CCG	2691 2700 GTG GGC CCG GGG GGC
Pro Pro Gly Ala	Pro Gly Ala	2ro Gly Ala	Pro Gly Pro	Vai Gly Pro Ala Gly
•	_	_	•	
2709 AAA AGC GGT GAT (	271E CGT GGT GAG	2727 ACC GGT CCG	2736 GCG GGC CCG	2745. 2754 CCC GTC GGC
				Ala Gly Pro Val Gly
2763 CCA GCG GGC GCC (	27.72 CGT GGC CCG	2781 GCC GGT CCG	2790 CAS GGC CCG	2799 2808 CCG GGT GAC AAA GGT
ero Ala Gly Ala	ard GTA Sto	Ala Gly Pro	Gin Gly Pro	Arg Gly Asp Lys Gly

FIG. 39D

		281	7		2826	5		2835	,		2344			2853	ı		2862
GAA	ACC	GG	CAA	CAC	GGC	GAC	CGI	. ccc	: AT	AAA 1	GGC	CAC	CC1	GCC	TTC	AGC	GGC
Glu	Th	GI	A CTO	GI	GT	ASP	Arg	l CT	, 116	e Lys	GLY	MIS	Arg	GTĀ	Yne	ser	CTA
		287	1		2880	)		2889	)		2898			2907			2916
CTG	CAC	GC.	r CCA	000	GGC	: 000	CCG	GGC	: AG	CCG	GGT	CAA	CAG	GGT	CCG	TCC	GGA
				·													
Leu	G1:	GL	y Pro	Pro	GI	, Pro	Pro	Gly	Se:	Pro	Gly	Glu	Gln	Gly	Pro	Ser	Gly
		292	5		2934	l		2943	:		2952			2961		2	2970
GCC	AGC	: cc	<del>,</del> cos	GCG	GGC	CCA	CGC	GGT	cc	CCG	GGC	AGC	GCG	GGT	GCG	$\infty$	GGC
Ala	Se	GL	/ Pro	Ala	Gly	Pro	Яrg	Gly	Pro	Pro	GŢĀ	Ser	Ala	GŢĀ	Ala	Pro	Gly
		2979	•		2988	1		2997			รถกร			3015		-	
AAA	ಆ	GG	CTG	AAC	GGT	CIG	CCG	GGC	cos	ATC	SSC	CCG	œ	GCC	CCA	ന്ദ	024
Lys	ysp	GL	Leu	Asn	Gly	Leu	Pro	Gly	6z3	Ile	GLY	Pro	\$r0	Gly	Pro	Arg	Gly
		3033	1		3042			3051			2050					_	
CGC	ACC	GGT	CAT	CCG	GGT	<b>cc</b>	GTG	CGT	œ	ထင	2000	CCG	ara <sup>s</sup>	3069	~~	3	078
Уrд	The	Gly	, Yzb	Ala	Gly	Pro	Val	Gly	Pro	Pro	Gly	Pro	2r0	Gly	Pro	Pro (	Glv
										3							
CCG	CCG	CST	ccc	ccc	AGC	GCG	GGT	TTC	620	TTC	7CC 7T4	<b>T-TV</b>	~~3	123	C) C	3	132
																J.G (	
PEO	250	GLy	Pro	bro	Ser	Ala	Gly	Phe	جو	Phe	Ser	Phe	Leu .	Pro	Gln I	Pro F	200
CAG	ಆರಿ	AAA	CCG	CAC	CAC	GGC	SCT.	722	TAC	TAC (	158						
													ś '				
GTU (	Glu	Lys	Ala	His.	Asp	Gly (	Gly :	Arg	Tyr	Tyr ;	1-g i	Ma					

FIG. 39E

EcoR1 start Oligo N1-1

GCGTGCCCCCCGATCGGTCCGACC-3'

3-GGCCCGGGCTACCCGGGCTCCCCGGGCCCACCGGGCCCCCCGGGGTCCAGCGGGGGCGAGCATTATTCGAACCC-5'

Oligo N1-2

EcoR1 Rsr II Oligo N1-3

3'-TACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTTTTTGCCGCTACTACCGCTTTGGCCCTTTGGCGCGTTATTGGAACCC-5'

Oligo N1-4

FIG. 40

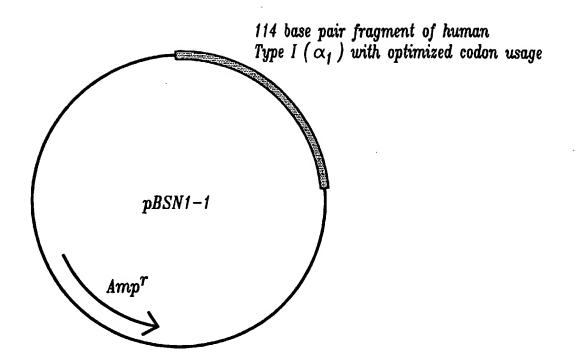


FIG. 41

۶٠	೧೬	CTG	9 AGC						27 AAA								ငၼ	54 GGC
	Gin	Leu	Ser	Tyr	Gly	Tyr	Asp	Glu	Lys	Ser	Thr	Gly	G1y	Ile	Ser	Vai	220	Gly
			63			72			81			90		•	99			108
	ಯ	ATG	GGT	CCG	AGC	GCC	$\infty$	CGT	GGC	CIG	$\infty$	GCC	$\alpha$ 3	CCA	337	ಯ	$\alpha$	GJT
	Pro	Met	Gly	Pro	Ser	Gly	Pro	Arg	Gly	Leu	Pro	Gly	623	Pro	Gly	Al a	Pro	G7À
	<b>∝</b> ⊊																	
	910																	

FIG. 42

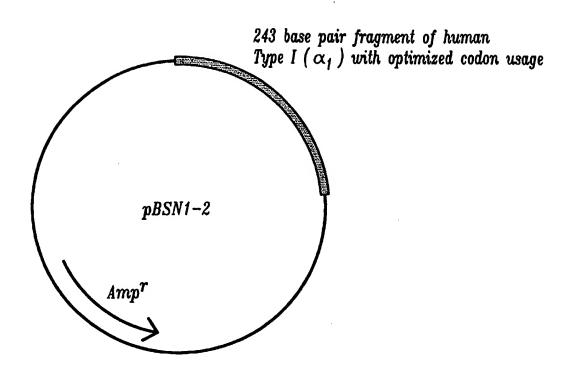


FIG. 43

		9				18			27			36			45			54		
5.	CAG	C:G	AGC	TAT	GCC	TAT	GAT	GAA	AAA	AGC	ACC	GGC	GGC	ATC	AGC	GTG	CCG	GGC		
	Gľv	Leu	Ser	Tyr	Gly	Tyr	Asp	Glu	Lys	Ser	Thr	Gly	Gly	Ile	Ser	Val	9.50	Gly		
			63			72			81			90			99			108		
	ന്നു	ATG															ccc	GGT		
	Pro	Met	Gly	Pro	Ser	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Pro	Pro	Gly	λla	Pro	Gly		
			117			126			135	•		144			153			162		
	ccs	CaG		TTT	CAG	GGT	ന്നു	CCG	GCC.	CAA	ccic	222	CDB	خک	CCT	~~	,	162		
														٠٠.	601		WC-			
	Pro	Gŗv	Gly	Phe	Gln	Gly	910	Pro	Gly	Glu	Pro	Gly	Glu	Pro	Gly	<u> 21</u> a	Ser	Gly		
			171		_	180			189			198			207			216		
	$\cos$	ATG	GGC	CCG	œc	GCC	CCG	$\infty$	GGT	œ	CCA.	GGC	AAA	220	507	CET	CLT	510		
	5to	んして	Gly	?ro	Arg	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Lys	As::	Gly	λsp	ÇZÁ	Gly		
																	_	-		
	GAA	$\alpha$		AAA			CCT	$\infty$												
	Glu	λla	GŢĀ	Lys	Pro	Gly	Arg	Pro												

FIG. 44

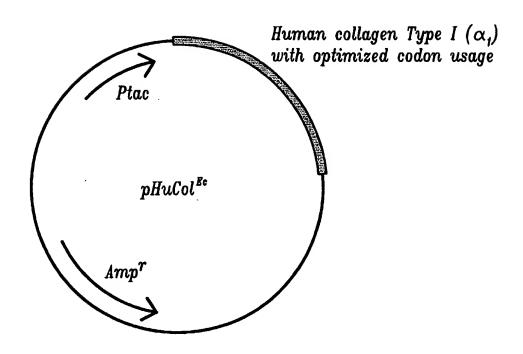


FIG. 45

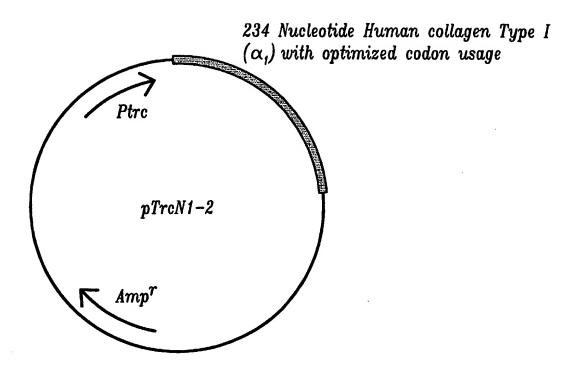


FIG. 46

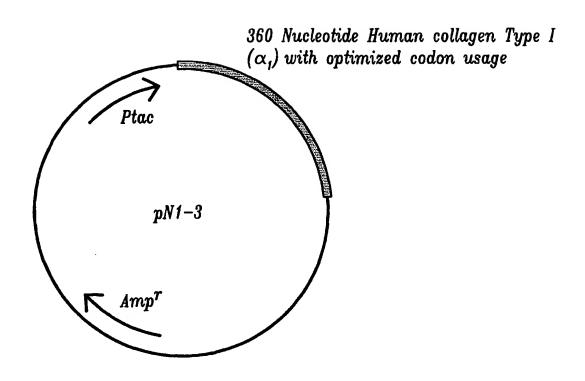


FIG. 47

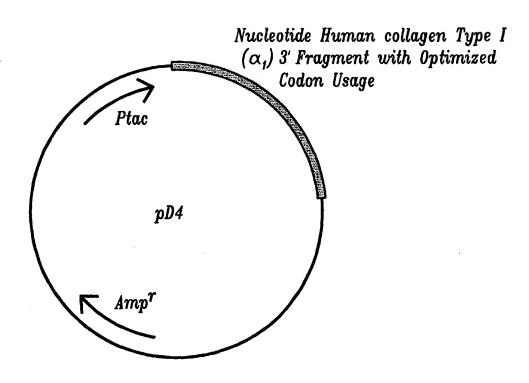


FIG. 48

9 18 - 27 36 45 54 5 CAG TAT CAT GGA MA GGA GTT GGA CTT GGC CCT GGA CCA ATG GGC TTA ATG GGA
Gln Tyr Asp Cly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu Met Gly
1
63 72 81 90 99 108 CCT AGA GGC CCA CCT GGT GCA GCT GGA GCC CCA GGC CCT CAA GGT TTC CAA GGA
Pro Arg Gly Pro Pro Gly Ala Ala Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly
117 126 135 144 153 162 OCT GCT GCT GCT GCT GCT GCT GCT GCT GCT G
Pro Ala Gly Glu Pro Gly Glu Pro Gly Gln Thr Gly Pro Ala Gly Ala Arg Gly
171 100 100 100
CCA GCT GGC CCT CCT GGC AAG GCT GGT GAA GAT GGT CAC CCT GGA AAA CCC GGA
Pro Ala Gly Pro Pro Gly Lys Ala Gly Glu Asp Gly His Pro Gly Lys Pro Gly
225 224 242 252
OCA OCT GGT GAG AGA GGA GTT GTT GGA OCA CAG GGT GCT CGT GGT TTC OCT GGA
Arg Pro Gly Glu Arg Gly Val Val Gly Pro Gln Gly Ala Arg Gly Phe Pro Gly
279 288 297 206 215
AND COLOR CITACT GGC TIC AAA GGC AIT AGG GGA CAC AAT GGT CTG GAT GGA
Thr Pro Gly Leu Pro Gly Phe Lys Gly Ile Arg Gly His Asn Gly Leu Asp Gly
TIG AAG GGA CAG GGT GCT GCT GGT GTG AAG GGT GAA GCT GGT GCC GCT GGT
Let the Giv Cin Dec Cin all
Leu Lys Gly Gln Pro Gly Ala Pro Gly Val Lys Gly Glu Pro Gly Ala Pro Gly
GAA AAT GGA ACT CCA GGT CAA ACA GGA GCC CGT GGG CTT CCT GGT GAG AGA GGA
Glu Asn Gly The Pro Gly Gln The Gly Ala Arg Gly Leu Pro Gly Glu Arg Gly
441 450 460
CGT GTT GGT GCC CCT GGC CCA GCT GGT GCC CGT GGC AGT GAT GGA AGT GTG GGT
Arg Val Gly Ala Pro Gly Pro Ala Gly Ala Arg Gly Ser Asp Gly Ser Val Gly
495 504 510
OCC GTG GGT CCT GGT CCC ATT GGG TCT GCT GGC CCT CCA GGC TTC CCA GGT
Pro Val Gly Pro Ala Gly Pro Ile Gly Ser Ala Gly Pro Pro Gly Phe Pro Gly
549 559 659
THE REPORT OF THE CONTROL OF THE CON
Ala Pro Gly Pro Lys Gly Glu Ile Gly Ala Val Gly Asn Ala Gly Pro Ala Gly
CCC GCC GGT CCC GGT GGA GTG GGT CTT CCA GGC CTC TCC GGC CCC GTT GGA
Pro Ala Gly Pro Arg Gly Glu Val Gly Leu Pro Gly Leu Ser Gly Pro Val Gly
657 666 675
CCT CCT GGT AAT CCT GGA GCA AAC GCC CTT ACT GGT GCC AAG GGT GCT GCT GGC
Pro Pro Gly Asn Pro Gly Ala Asn Gly Leu Tar Gly Ala Lys Gly Ala Ala Gly
tad all oily hid Aid Gly

FIG. 49A

		-				720			729			738			747		~~	756
~~	$\sim$	G	CC I	GTT	CT	œ	CCT	$\infty$	GGC	CTC	$\alpha$	GGA	$\infty$	œc	GGT	ATT	CCI	
		-	~-								Dro.	Gly	Pro	Arg	Gly	Ile.	Pro	Gly
Leu	Pro	G	Ίy	Val	Ala	Gly	Ala	Pro	GIĀ	JEU.	210	ULJ		3	•			
		_	166			774			783			`792			801	C) C	~~	810
~~	CT1	٠ (	GT	GCT	œ	GGT	GCT	ACT	GGT	$\infty$	AGA	GC3/	CTT	GTT	GGI	GAG	<u>ur</u>	661
										B1 2	7~~	Gly	Teu	Val	Gly	Glu	Pro	Gly
Pro	٧al	. (	Gly	Ala	Ala	Gly	YIZ	The	GTÅ	ALA	ALY	GIJ			Gly			
		,	010			828			837			846			855		~~	864
CCA	GC	. (	GGC	TCC	AAA	GGA	GAG	AGC	GGT	AAC	AAG	GGT	GAG	ccc	GGC	TCT	GCT	
		-								7	Tag	Glv	Glu	Pro	Gly	Ser	Ala	Gly
Pro	Ala	9	Gly	Ser	Lys	GTA	GLU	Ser	GIŞ	A3!!	Lys	049	<b>V</b>					Gly
			873			892			891			900			909			918
$\alpha$	CA	A	GGT	CI	cor	GGI	$\infty$	AGT	GGT	GAA	, GA	GGA	AAG	ALA	GGC	<u> </u>	AAL	GGG
																		Gly
Pro	G1	n	GTĀ	¥1C	Pro	GTA	220			-		,		-				
			927			936	5		945			954			963		~	972
GA	A GC	T	GGA	TC	c ex	GGC	: œ	· 002	CGA	, CC1	. 007	r GG	CR	ALSA 				GGT
	. hi	_	Glu	Se	- Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Leu	Arg	Gly	Ser	Pro	Gly
GI	י אי	۵,	Gry				-											
			981			990	)		999	)	~	1008	· ~~	• <b>እ</b> ጥ	1017		• ~~	1026 CCT
TC	rα	T	GGI	CT	r œ	GG	i GC	GA		. Alar								GGT
Se	- Ar	<u> </u>	Gly	Le	ı Pro	Gly	/ Ala	As	Gly	Arg	Ala	a Gly	val	. Met	: Gly	Pro	Pro	Gly
.~	• ~		.035		1 BČ1	1044			1053 (GC)		·	1062		רבב י	1071 (GC)			1080 GGT
		_						· <del></del>										
Se	r Ar	g	Gly	Ala	a Sez	Gl <sub>3</sub>	Pro	Ala	Gly	· Val	Arc	gly	Pro	Asn	(Gly	Asp	Ala	Gly
		1	.089			1098	,		1107	,		1116			1125			1134
œ	: α				; œ			ATC						CI				GGA
		-								-	-							
Αr	J Pr	0	Gly	Glu	ı Pro	Gly	Leu	Met	Gly	Pro	Arg	Gly	Leu	Pro	GLY	Ser	Pro	Gly
		1	143			1152	2		1161			1170	1		1179		٠	1188
AA	TA 1	c	GGC	$\alpha$	CI	G	L AAA	GAA	GGT	CI	GIX	: GGC	CTC	CI	GCC	ATC	GAC	GGC
		-					. 7			·	17-1			D				
ASI	; TT	е	GIĀ	PIC	AL	GTA	гÃа	GIU	GTĀ	PIC	· val	. GLY	Teu	PEO	GIY	ще	ASP	Gly
		1	197			1206	5		1215			1224			1233			1242
AG	$\infty$	T	GGC	$\infty$	TTA A	. eec	: CCA	GCI	GCA	. GCCA	AGA	GCA	GAG	CCI	GGC	AAC	ATT	GGA
A		~	Glu	Dro	Tie	Glu	Pm	Ala	Glu	Ala	Am	. <u></u>	Glu	Pro	Glv	760	Tle	Gly
		•	Gry	210	, 440			,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	· OLy			,,	920		, Gra	7511	110	GLY
:			251			1260			1269			1278			1287			1296
TI	: α	T	GGA	. α	: AAA	630	: 000	ACT	GGI	GAI	· œ	· 6600	AAA	AAC	GGT	GAT	AAA	GGT
Phe	Pr	0	Glv	Pro	Ivs	Glv	Pro	Thr	Glv	Asp	Pro	Glv	Lvs	Asn	Glv	Asp	Lvs	Gly
		-	1		-2								-		1		_3 -	1
			305		. ~~-	1314			1323		· ~~-	1332			1341			1350
-A.	. GC	7		CTI			GC1	. <del></del>		. <del></del>	. —		ur 	GAT		AAC	AAT	GGT
His	: Al	a	СГĀ	Lei	a Ala	Gly	Ala	Arg	Gly	Ala	Pro	Gly	Pro	Asp	Gly	Asn	Asn	Gly
ري-	בר י		359 GGA		· CCT	1368 : GG2			1377 : GGT		CAN	1386 GGT			1395 . GGT	44D		GGT
		_																

FIG. 49B

Ala Gln Gly Pro Pro Gly Pro Gln Gly Val Gln Gly Gly Lys Gly Glu Gln Gly 1449 1440 COO GAT GET COT COA GET THE CAG GET CHE COT GET COX TOA GET COX GET GET Pro Asp Gly Pro Pro Gly Phe Gln Gly Leu Pro Gly Pro Ser Gly Pro Ala Gly 1503 1494 GAA GTT GCC AAA CCA GGA GAA AGG GGT CTC CAT GGT GAG TTT GGT CTC CCT GGT 1485 Glu Val Gly Lys Pro Gly Glu Arg Gly Leu His Gly Glu Phe Gly Leu Pro Gly 1548 CCT GCT GGT CCA AGA GGG GAA CGC GGT CCC CCA GGT GAG AGT GGT GCT GCC GGT 1530 Pro Ala Gly Pro Arg Gly Glu Arg Gly Pro Pro Gly Glu Ser Gly Ala Ala Gly 1611 1593 1602 OCT ACT GGT OCT ATT GGA AGC OGA GGT OCT TOT GGA COC OCA GGG OCT GAT GGA Pro Thr Gly Pro Ile Gly Ser Arg Gly Pro Ser Gly Pro Pro Gly Pro Asp Gly 1647 . 1665 1638 AME AME GET GAM CET GET GTG GTT GET GET GET GEC MET GET GET CEM TET GET Asn Lys Gly Glu Pro Gly Val Val Gly Ala Val Gly Thr Ala Gly Pro Ser Gly 1692 1701 CCT AGT GGA CTC CCA GGA GAG AGG GGT CCT CCT GGC ATA CCT GGA GGC AAG GGA Pro Ser Cly Leu Pro Gly Glu Arg Gly Ala Ala Gly Ile Pro Gly Gly Lys Gly 1746 1755 CAA AAG GGT GAA OCT GGT CTC AGA GGT GAA ATT GGT AAC OCT GGC AGA GAT GGT Glu Lys Gly Glu Pro Gly Leu Arg Gly Glu Ile Gly Asn Pro Gly Arg Asp Gly 1809 GCT OGT GGT GCT CAT GGT GCT GTA GGT GCC GCT GGT GCT GGA GCC ACA GGT Ala Arg Gly Ala His Gly Ala Val Gly Ala Pro Gly Pro Ala Gly Ala Thr Gly 1854 1863 Asp Arg Gly Glu Ala Gly Ala Ala Gly Pro Ala Gly Pro Ala Gly Pro Arg Gly AGC OCT GGT GAA OGT GGC GAG GTC GGT OCT GCC GGC CCC AAC GGA TIT GCT GGT Ser Pro Gly Glu Arg Gly Glu Val Gly Pro Ala Gly Pro Asn Gly Phe Ala Gly 1962 1971 1980 COG GCT GGT GCT GGT CAA COG GGT GCT AAA GGA GAA AGA GGA GCC AAA GGG Pro Ala Gly Ala Ala Gly Gln Pro Gly Ala Lys Gly Glu Arg Gly Ala Lys Gly 2016 2025 2034 OCT AAG GGT GAA AAC GGT GTT GTT GGT OOC ACA GGC OOC GTT GGA GCT GCT GGC Pro Lys Gly Glu Asn Gly Val Val Gly Pro Thr Gly Pro Val Gly Ala Ala Gly 2079 2088 2097 CON NNN GGT CCA AAT GGT CCC CCC GGT CCT GCT GGA AGT CGT GGT GAT GGA GGC

FIG. 49C

Pro	Xxx	Gly	Pro	Asn	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ser	Arg	Gly	ÇZA	Gly G	ly
		2115	3 TYC		2124		~	2133	·~~	~	2142			2151	~~	CCA GO	50
Pro	Pro	Gly	Met	Thr	Gly	Phe	Pro	Gly	Ala	Ala	Gly	Arg	Thr	Gly	Pro	Pro G	ly
		2169			2178	~~		2187	~~~		2196			2205		221	
<u> </u>	TCT	GGT	ATT	TCT	GGC	<del></del>	<del>ст</del>	GGT	<del></del>	<u>~~</u>	GGT	CCT	GCT	GGG	AAA	 Gyy C	3G
Pro	Ser	Gly	Ile	Ser	Gly	Pro	Pro	Gly	Pro	Pro	Сĵå	Pro	Ala	Gly	Lys	Glu G	ly
		2223			2232	<b>63.6</b>		2241			2250			2259	<i>-</i>	220	
CII	CGT	GGA	CCN	CZ4		GAC.	CAA		CCA	GCA	GGC	CCA		GGA	G-2.A	GTA G	Ξλ Ξλ
Leu	Arg	Gly	Pro	Arg	Gly	czÁ	Gln	Gly	Pro	Ala	Gly	Arg	Pro	Gly	Glu	Val G	ly
		2277	~~~		2286	~~~		2295	~~		2304	~~~		2313	c> c	232	
GCA		GGI						901								<u>ccr</u> c	 31
Ala	Pro	Gly	Pro	Pro	Gly	Phe	Ala	Gly	Glu	Lys	Gly	Pro	Ser	Gly	Glu	Ala G	Ly
9		2331			2340			2349						2367	~~~	237	
ACT	GCT	GGA	<u>~~</u>	<u>СТ</u>	- <del></del>	ACT	CCA	GGT	<del>сст</del>	CAG	GGT	CTT	CIT	GGT	GCT	<u></u>	JT 
Thr	Ala	Gly	Pro	Pro	Gly	The	Pro	Gly	Pro	Gln	Gly	Leu	Leu	Gly	Ala	Pro G	ŗy
		2395			2394			2403			2412			2421		243	
ATT	CTG	GGT	CIC	Œī	<u>eee</u>	TCG	yCy	GGT	G574	<b>©</b> ∓	<u> </u>	CIA	CCT	GGT	GIT	<u>ect</u> e	3T 
Ile	Leu	Gly	Leu	Pro	Gly	Ser	Arg	Gly	Glu	Arg	СЈĀ	Leu	Pro	elà	Val	Ala G	ly
		2439			2448			2457			2466			2475		248	
CCT	GTG	2439 GGT	GAA									œ <sub>T</sub>			<u></u>	CGT G	
	GTG	GGT		<u>CT</u>	GGT	CCT	CTT	GGC	ATT ——	<u>~~</u>	<u></u>		CCT	GGG			37
Ala	GTG Val	GGT Gly 2493	Glu	Pro	GGT Gly 2502	CCT Pro	Leu	GGC Gly 2511	ATT —— Lle	Ala	GGC Gly 2520	Pro	CCT Pro	GGG  Gly 2529	Ala	CGT GG Arg G	ly 38
Ala	GTG Val	GGT Gly 2493	Glu	Pro	GGT Gly 2502	CCT Pro	Leu	GGC Gly 2511	ATT —— Lle	Ala	GGC Gly 2520	Pro	CCT Pro	GGG  Gly 2529	Ala	CGT GG Arg G	ly 38
Ala	Val	Gly 2493 GGT	Glu	Pro	GGT Gly 2502 GGT	Pro	Leu	GGC Gly 2511 GGA	ATT Lle	Ala	GC Gly 2520 GGT	Pro	Pro	GGG Gly 2529 GGT	Ala	CGT GG Arg G	37 Ly 38 37
Ala CCT Pro	CCT Pro	GGT Gly 2493 GGT Gly	Glu GCT  Ala	CTG Pro	GGT Gly 2502 GGT Gly 2556	Pro AGT	CCT Leu CCT Pro	GGC Gly 2511 GGA Gly 2565	Ile GIC Val	Ala AAC Asn	GGC Gly 2520 GGT Gly 2574	Pro GCT Ala	CCT Pro	GGG Gly 2529 GGT Gly 2583	GAA GLu	CGT GC Arg G CCT GC Ala G	37 Ly 38 37 Ly
Ala CCT Pro	CCT Pro	GGT Gly 2493 GGT Gly	Glu GCT  Ala	CTG Pro	GGT Gly 2502 GGT Gly 2556	Pro AGT	CCT Leu CCT Pro	GGC Gly 2511 GGA Gly 2565	Ile GIC Val	Ala AAC Asn	GGC Gly 2520 GGT Gly 2574	Pro GCT Ala	CCT Pro	GGG Gly 2529 GGT Gly 2583	GAA GLu	CGT GC Arg G CCT GC Ala G	37 Ly 38 37 Ly
Ala CCT Pro	CCT Pro	GGT Gly 2493 GGT Gly 2547	GCT Ala	CTG Pro	GGT Gly 2502 GGT Gly 2556 GGG	Pro AGT Ser	CCT Leu CCT Pro	GGC Gly 2511 GGA Gly 2565 GGT	GIC Val	Ala AAC Asn	GGC Gly 2520 GGT Gly 2574 GGT	Pro GCT Ala	CCT Pro CCT Pro	GGG Gly 2529 GGT Gly 2583 GGT	GAA Glu CAA	CGT GC Arg G CCT GC Ala G	37 1y 38 37 1y 92 3A
CCT Pro	CCT Pro	GGT Gly 2493 GGT Gly 2547 GGC GGC	GCT Ala AAC ASn	CT Pro GIG Val CT Pro	GGT Gly 2502 GGT Gly 2556 GGS Gly 2610	AGT Ser AAC	CCT Leu CCT Pro	GCC Gly 2511 GCA Gly 2565 GCT Gly 2619	GIC Val	AAC Asn CCA Pro	2520 Gly 2520 GGT Gly 2574 GGT Gly 2628	Pro GCT Ala CGC Arg	Pro  GAT  Asp	GGG Gly 2529 GGT Gly 2583 GGT Gly 2637	GAA Glu CAA Gln	25: CCT CC CCT CC CCT CC CCT CC CCT CC CCT CC CC	38 37 11 11 11 12 38 37 11 12 46
CCT Pro	CCT Pro	GGT Gly 2493 GGT Gly 2547 GGC GGC	GCT Ala AAC ASn	CT Pro GIG Val CT Pro	GGT Gly 2502 GGT Gly 2556 GGS Gly 2610	AGT Ser AAC	CCT Leu CCT Pro	GCC Gly 2511 GCA Gly 2565 GCT Gly 2619	GIC Val	AAC Asn CCA Pro	2520 Gly 2520 GGT Gly 2574 GGT Gly 2628	Pro GCT Ala CGC Arg	Pro  GAT  Asp	GGG Gly 2529 GGT Gly 2583 GGT Gly 2637	GAA Glu CAA Gln	CGT GC Arg G  25: GCT GC Ala G  25: CC GC Pro G	38 37 11 11 11 12 38 37 11 12 46
Ala CCT Pro CGT Arg	CCT Pro	GGT Gly 2493 GGT Gly 2547 GGC GGA 2601 GGA	GCT Ala AAC Asn	GTG Pro	2502 Gly 2502 GGT Gly 2556 GGG Gly 2610	AGT Ser AAC ASn	CCT Leu CCT Pro GAT Asp	GCC Gly 2511 GCA Gly 2565 GGT Gly 2619	GIC Val	ALCA AST	2520 Gly 2520 GGT Gly 2574 GGT Gly 2628 GGT	Pro  GCT Ala  CGC Arg	CCT Pro CCT Pro GAT Asp	2529 GGT Gly 2583 GGT Gly 2583 GGT GGT	GAA Glu CAA Gln	25: CCT CC CCT CC CCT CC CCT CC CCT CC CCT CC CC	38 37 11y 92 33 11y 46 37
Ala CCT Pro CGT Arg	CCT Val	2493 GIy 2493 GGT GIY 2547 CGC GIY 2601 GGA	GCT Ala  AAC Asn  GRG GRG	CTG Pro Val  CTG Pro	CGT	Pro AGT Ser AAC TAC Tyr	CCTT Leu CCTT Pro GAT Asp	G1y 2511 G2y 2515 G1y 2565 G1y 2619 G1y 2619	GIC Val	Ala  AAC  Asn  CCA  Pro	GGC GIY  2520 GGT GIY  2574 GGT GIY  2628 GGT GIY  2628 GGT GIY	Pro  GCT  Ala  GGC  Arg  CCC  Pro	CCT Pro CCT Pro GAT Asp	GGG Gly 2529 GGT Gly 2583 GGT Gly 2637 GGT Gly 2637 GGT G1y 2637	GAA Glu CAA Gln GCT	CGT GC Arg G  25: GCT GC Ala G  25: GCT GC Ala G  25: GCA GC Ala G  Ala G  27:	38 37 14 38 37 14 32 34 46 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 14 14 14 14 14 14 14 14 14 14 14 14
Ala CCT Pro CGT Arg	CCT Val	2493 GIy 2493 GGT GIY 2547 CGC GIY 2601 GGA	GCT Ala  AAC Asn  GRG GRG	CTG Pro Val  CTG Pro	CGT	Pro AGT Ser AAC TAC Tyr	CCTT Leu CCTT Pro GAT Asp	G1y 2511 G2y 2515 G1y 2565 G1y 2619 G1y 2619	GIC Val	Ala  AAC  Asn  CCA  Pro	GGC GIY  2520 GGT GIY  2574 GGT GIY  2628 GGT GIY  2628 GGT GIY	Pro  GCT  Ala  GGC  Arg  CCC  Pro	CCT Pro CCT Pro GAT Asp	GGG Gly 2529 GGT Gly 2583 GGT Gly 2637 GGT Gly 2637 GGT G1y 2637	GAA Glu CAA Gln GCT	25: CCT GC Ala G. Pro G. Pro G. Ala G	38 37 14 38 37 14 32 34 46 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 37 14 14 14 14 14 14 14 14 14 14 14 14 14
Ala CCT Pro CGT Arg CAC	CCT Pro	GGT Gly 2493 GGT GGLY GGC GGLY GGC GGC GGLY GGC GGC GGC GGC GGC GGC GGC GGC GGC GG	GCT Ala AAC Asn GAG Glu	CTG Pro Val  CTG Pro CCC Arg	2502 2502 2502 Gly 2556 Gly 2610 Gly 2664 Gly	AGT AGT ASD TAC	CCTT Leu  CCTT Pro  GAT Asp  CCTT Pro	GSC Gly 25615 GGY 2565 GGT Gly GGY 2619 GGC GGT GGT GGT GGT GGT GGT GGT GGT GGT	GTC Val CCC Pro	ALAC ASR Pro	CGC Gly 2574 GGT Gly 2628 GGT Gly 2682 GGT Gly 2682 GGT Gly 2682 GGT GGT GGT GGT GGT GGT GGT GGT GGT GG	CCC ALTY	CCT Pro  CAT Asp CAT Val	GGG GIY 2529 GGT GIY 2583 GGT GIY 2637 GGY GGY 61 GIY 2637 GGY 61 GIY	GAA GLu CAA GLI GLI GLI AL AAC	CGT GC Arg G  25: GCT GC Ala G  25: GCT GC Ala G  25: GCA GC Ala G  Ala G  27:	38 37 1y 38 37 1y 46 37 1y 00 37
Ala CCT Pro	CCT Pro GAT Asp Lys	GGT Gly 2493 GGT GGT GGT GGT GGT GGT GGT GGT GGT GG	Glu GCT Ala AAC Asn GRG Glu CCT	CCT Pro CCC Arg	CGT Gly 2502 25566 GGT Gly 2610 GGT Gly 2664 GGT Gly 2718	AGT AGT ASD TAC	CCTT Leu CCTT Pro GAT Asp CCTT Pro GTG Val	GGC Gly 2511 GGA GGY 2565 GGT GGY GGY 2619 GGC GGT GGY 2727 GGY 2727 GGY 2727 GGY 2727 GGY 2727	GTC Val	AAC Asn CCA Pro ATT Ile CCT Ala	GCC Gly 25200 GGT Gly 2574 GGT Gly 2628 GGT Gly 26882 GGC GGT Gly 2736	Pro GCT Ala CCC Arg Pro	CCT Pro  CCT Pro  GAT Asp  CAT CAT	GGG G1y 2529 2529 GGT G1y 2583 GGT G1y 2637 G1y 2637 G1y 2691 GGA GGA GGA G1y 2745	CAA Glu CAA Gln GCT Ala AAC Asn	CGT GC Arg G:  25: GCT GC Ala G:  Pro G:  26: GCA GC Ala G:  Ala G:  Arg G:  Arg G:  Arg G:  Arg G:	38 37 1y 38 37 1y 46 37 1y 46 37 1y 54
Ala CCT Pro	CCT Pro GAT Asp Lys	GGT Gly 2493 GGT GGT GGT GGT GGT GGT GGT GGT GGT GG	Glu GCT Ala AAC Asn GRG Glu CCT	CCT Pro CCC Arg	CGT Gly 2502 25566 GGT Gly 2610 GGT Gly 2664 GGT Gly 2718	AGT AGT ASD TAC	CCTT Leu CCTT Pro GAT Asp CCTT Pro GTG Val	GGC Gly 2511 GGA GGY 2565 GGT GGY GGY 2619 GGC GGT GGY 2727 GGY 2727 GGY 2727 GGY 2727 GGY 2727	GTC Val	AAC Asn CCA Pro ATT Ile CCT Ala	GCC Gly 25200 GGT Gly 2574 GGT Gly 2628 GGT Gly 26882 GGC GGT Gly 2736	Pro GCT Ala CCC Arg Pro	CCT Pro  CCT Pro  GAT Asp  CAT CAT	GGG G1y 2529 2529 GGT G1y 2583 GGT G1y 2637 G1y 2637 G1y 2691 GGA GGA GGA G1y 2745	CAA Glu CAA Gln GCT Ala AAC Asn	25: CCT GC GCA GCA GCA GCA GCA GCA GCA GCA GCA	38 37 1y 38 37 1y 46 37 1y 46 37 1y 54
Ala CCT Pro CGT Arg CAC His	CCT Pro GAT Asp Lys CCT Pro	GGT Gly 2493 GGT GGT GGG GGG GGG GGG GGG GGG GGG GG	GAG GAG CCT Ala AAC Asn GAG CT Pro	CCT Pro Val CCT Pro CCC Arg	CGT	AGT Ser AAC TAC CCC Pro	CCT Pro  GAT Pro  GTG Val	GSC Gly 2511 Gly 2565 Gly GCC Gly GCC Gly GCC GCC GCC GCC GCC GCC GCC GCC GCC GC	ATT Lie GIC Val CCC Pro AAT Asn CCT CT CCT	ATT The CCT Ala	GCC Gly 2520 GGI Gly 2574 GGI Gly 2628 GGI Gly 2682 GGI Gly 27366 GGI 2736 GGI 2736 GGI 2736 GGI 2736 GGI 2736 GGI 2736	Pro  GCT Ala  GGC Arg  CCC Pro  AAA  CGC CCT	CCT Pro  CCT Pro  GAT Asp  CAT  CAT  CAT  GTT  CAT	GGG G1y 2529 GGT G1y 2583 GGT G1y 2637 G1y 2691 GGA GGA G1y 2745 GGY 2745	GAA GLu CAA GLn GCT Ala AAC Asn	25: CCT CCT CCT CCT CCT CCT CCT CCT CCT CC	38 37 1y 38 37 1y 46 37 1y 00 37 1y 54 37 1y
Ala CCT Pro CGT Arg CAC His	CCT Pro GAT Asp Lys CCT Pro	GGT Gly 2493 GGT GGT GGG GGG GGG GGG GGG GGG GGG GG	GAG GAG CCT Pro	CCT Pro  GIG CT Pro  CCT Pro  CCT Pro  CCT CCT CCT CCT CCT CCT CCT CCT CCT C	CGT	AGT Ser AAC TAC CCC Pro	CCT Pro  GAT Pro  GTG Val	GSC Gly 2511 Gly 2565 Gly GCC Gly GCC Gly GCC GCC GCC GCC GCC GCC GCC GCC GCC GC	GIC Val CCT Asn CCT Pro	AAC ASn Pro ATT Ile GCT Ala	GCC Gly 2520 GGI Gly 2574 GGI Gly 2628 GGI Gly 2682 GGI Gly 27366 GGI 2736 GGI 2736 GGI 2736 GGI 2736 GGI 2736 GGI 2736	GCT Ala CCC Arg Pro Lys GCT Ala	CCT Pro  CCT Asp  GTT Val  CAT His	GGG G1y 2529 GGT G1y 2583 GGT G1y 2637 G1y 2691 GGA GGA G1y 2745 GGY 2745	GAA Glu CAA Gln GCT Ala AAC Asn	CGT GC Arg G:  25: GCT GC Ala G:  Pro G:  26: GCA GC Ala G:  Ala G:  Arg G:  Arg G:  Arg G:  Arg G:	38 37

FIG. 49D

$\alpha$	AGT	GGC	CCA	CAA	GGC	ATT	ŒΤ	GGC	GAT	AAG	GGA	GAG	$\infty$	GGT	GAA	AAG	GGG
Pm	Ser	Glv	Pro	Gla	Glv	Tle	Arri	Glv	Asp	LVS	Glv	Glu	Pro	Glv	Glu	T.vs	Gly
110		023			ردن		,	<b>U</b>	۲.	٠,٠	<b>0-</b> 3			<b>U</b> _j		<b>~</b> , 5	OLY
		2817			2826									2853			2862
$\infty$	AGA	GGT	CTT	$\alpha$	œ	TTC	AAG	GCY.	CAC	AAT	GGA	TTG	CAA	GT	CIG	$\infty$ r	GGT
P==	A	Cly	Ten	Pro	Gly	Dhe	Lve	Gly	Wie.	Asn	Gly	Ten	Gla	Gly	Leu	P	C1
510	λtg	Gry	Deu	110	GTÅ		2,3	043	.43	~	GL.y	2	un.	GLY	LEG	FIO	GTÅ
	;	2871		:	2880		:	2889			2898		:	2907		:	2916
ATC	GCT	GGT	CAC	CAT	CCI	CAT	CAA	GGT	CCT	$\alpha$	GCC	TCC	GIG	GGT	$\alpha$	CCT	GGT
		<u></u>			<u></u>	3		C)	31-	Desc	<u></u>		17-1				
TTE	ALA	GIŞ	HL.S	MLS.	CTĀ	ASP	GIII	GTÅ	MI	PIO	GTĀ	ser	Val	CTÀ	Pro	ALA	GTÅ
	:	2925		:	2934		:	2943		:	2952		1	2961		:	2970
CCT	AGG	GGC	$\alpha$												$\alpha$ c		GGA
Pro	Arg	Gly	Pro	Ala	GIA	Pro	Ser	Gly	Pro	Ala	Gly	Lys	Asp	GIŢ	Arg	Thr	Gly
	:	2979		:	2988		:	2997			3006		3	3015		3	3024
CAT	$\infty$ T	GGT	ACG	GTT	GGA	$\alpha$	CCT	GGC	ATT	ŒA	GCC	$\alpha$	CAG	GGT	CAC		
HIS	Pro	GIA	Thr	Vai	CIA	Pro	ALA	GLY	Пe	Arg	GTA	Pro	.GLn	GTĀ	His	Glu	GJA
		3033		:	3042			3051		:	3060		3	3069		3	1078
$\alpha$	GCI	GĠC	$\infty$	$\alpha$	GGT	$\infty$	$\infty$ T	GGC	$\infty$ T	CTT	œ	$\alpha$	CTA	GGT	GTA	AGC	GGT
		~~~	D						~								
PIO	ALG	GIY	PIO	PIO	GTĀ	Pro	Pro	crā	Pro	Ten	GTA	Pro	Leu	Gly	Val	Ser	Gly
	:	3087		3	3096		3	105		3	3114						
GGT	GGT	TAT	CAC	TTT	<b>GGT</b>	TAC	GAT	GGA	GAC	TTC	TAC	AGG	GCT	3'			
		~															
era	GTA	Tyr	qzA	hue	GTÅ	ıyr	Asp	GTA	Asp	Phe	Tyr	yra	Ala				

FIG. 49E

5'	CAG	TAC	9 GAC	CCT	ĄĄĄ	18 GGC	gta	GGC	27 CTG	GGT	CCG	36 GGT	ccc	ATG	45 GGC	CTG	ATG	54 GGT
				Gly														
				CCA  Pro														
	CCG	GCG	117 GGT	GAA	ccc	126 GGC	GAA	CCG	135 GGT	CAG	ACG	144 GGT	CCG	GCG	153 GGT	GCT	CGC	162
	CCG	GCT	171 GGC	CCA	CCG	180 GGC	AAA	GCT	189 GGC	GAA	GAC	198 GGT	CAC	CCG	207 GGT	AAG	CCA	216
	ccc	ccg	225 GGC	GAA	CGI	234 GGC	GTC	GIG	243 GGT	CCG	CAA	252 GGT	GCG	CGT	261 GGT	TTC	CCG	270
				CTG  Leu	·											<u></u>		324 GGT  Gly
	·			CAA  Gln														
				ACG Thr														
				GCT  Ala														
	CCG	GTT	495 GGC	CCT  Pro	CCC	504 GGT	CCG	ATT	513 GGT	TCC	GCT	522 GGC	CCT	ccc	531 GGT	TTC	ccc	540 GGT
	GCG	CCG	549 GGT	CCG  Pro	AAG	558 GGT	GAG	ATC	567 GGC	GCG	GTT	576 GGC	AAC	GCA	585 GGC	CCG	GCT	594 GGT
	CCA	GCC	603 GGC	CCT  Pro	CGT	612 GGC	GAA	GTC	621 GGT	CTG	ccc	630 GGT	CTG	AGC	639 GGT	ccc	GTA	648 GGC

		657			666			675	<b>~</b>	100	684	003	222	693	ccc	COT	702
			AAC														
Pro	Pro	Gly	Asn	Pro	Gly	Ala	Asn	Gly	Leu	Thr	Gly	Ala	Lys	Gly	Ala	Ala	Gly
		711			720			729			738			747	\mm	~~	756
			GTT														
Leu	Pro	Gly	Val	Ala	Gly	Ala	Pro	Gly	Leu	Pro	Gly	Pro	Arg	Gly	Ile	Pro	Gly
		765			774			783			792			801			810
CCG	GTA	GGC	GCA	GCC	GGT	GCA	ACT	GCT	GCC.	CGT	GGC	CTG	GTT	GGC	GAA	ccc	GGT
Pro	Val	Gly	Ala	Ala	Gly	Ala	Thr	Gly	Ala	Arg	Gly	Leu	Val	Gly	Glu	Pro	Gly
		819			828			837			846			855			864
CCG	GCG	GGT	TCT	AAA	GGC	GAA	AGC	CCT	AAC	AAA	GGT	GAG	CCG	GGT.	TCC	GCG	GGC
Pro	Ala	Gly	Ser	Lys	Gly	Glu	Ser	Gly	Asn	Lys	Gly	Glu	Pro	Gly	Ser	Ala	Gly
		873			882			891			900			909			918
ccc	CAG	GGT	CCG	CCG		CCG	AGC			GAA	GGT	AAA	CGT	GGT	ccc	AAC	
Pro	Gln	Gly	Pro	Pro	Gly	Pro	Ser	Gly	Glu	Glu	Gly	Lys	Arg	Gly	Pro	Asn	Gly
		927		,.	936			945			954			963			972
GAG	GCT		TCC	GCA			ccc			$\infty$			CCT		AGC	$\infty$	
Glu	Ala	Gly	Ser	Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Leu	Arg	Gly	Ser	Pro	Gly
		981			990			999			1008			1017		,	1026
AÇC	CGT		CTG	CCG			GAC										
Ser	Arg	Gly	Leu	Pro	Gly	Ala	Asp	Gly	Arg	Ala	Gly	Val	Met	Gly	Pro	Pro	Gly
		1035			1044			1053			1062		;	1071			1080
TCC	CGT	GGT	GCC	TCT	GGT	CCG	GCT	GGT	GTC	CCI	GGT	CCG	AAT	GGC	GAC		
Ser	Arg	Gly	Ala	Ser	Gly	Pro	Ala	Gly	Val	Arg	Gly	Pro	Asn	Gly	Asp	Ala	Gly
		1089			1098			1107		:	1116			L125		,	L134
CGT	CCG	GGT	GAA	CCG	GCC	CIG	ATG	GGT	CCC	CCT	GGC	CTG			AGC	CCG	GGT
Arg	Pro	Gly	Glu	Pro	Gly	Leu	Met	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Ser	Pro	Gly
		1143			1152			1161		•	1170			1179		•	L188
AAC	ATT	CCT	CCC	GCG	GGT	AAG	GAG	CCT	ccc			CTG			ATT		
Asn	Ile	Gly	Pro	Ala	Gly	Lys	Glu	Gly	Pro	Val	Gly	Leu	Pro	Gly	Ile	Asp	Gly
	:	1197			1206			1215			1224			1233	·	•	L242
CGT	CCG	GGT	CCG	ATC	GGC	CCT			GCT			GAG	ccc	GGT	AAC	ATC	GGT
Arg	Pro	Gly	Pro	Ile	Gly	Pro	Ala	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Asn	Ile	Gly
		1251			1260			1269			1278			1287		1	- 1296
TTT	CCG	GGT	CCG			CCG			GAC	ccc	GGC	AAG	AAC	GGT	CAT	AAA	GGC
Phe	Pro	Gly	Pro	Lys	Gly	Pro	Thr	Gly	Asp	Pro	Gly	Ĺys	Asn	Gly	Asp	Lys	Gly
		1305			1314			1323			1332		1	341		1	1350
CAT	GCA	GCT	CIG	GCA	CCT	GCC	CGT	GGT	GCA	CCC	CCT	222	GAT	GGT	AAC	AAT	GGT
Hie	Ala	Gly	Leu	Ala	Gly	Ala	Arg	Gly	Ala	Pro	Gly	Pro	Asp	Gly	Asn	Asn	Glv

1359 GCG CAG GGT								1395 AAA GGT	1404 GAA CAG OGT
Ala Gln Gly	Pro Pro	Gly F	ro Gln	Gly	Val	Gln Gly	Gly	Lys Gly	Glu Gln Gly
			TC CAG		CTG		ccc		1458 CCG GCT GGT
Pro Ala Gly	Pro Pro	o Gly I	he Gln	Gly	Leu	Pro Gly	Pro	Ser Gly	Pro Ala Gly
		- <b></b> -	EAA CGT		CTC.	CAT GGC	GAG		CTG CCG GGT
Glu Val Gly	Lys Pro	o Gly (	Slu Arg	Gly	Leu	His Gly	Glu	Phe Gly	Leu Pro Gly
	cce ce		GAG CGC		CCT	CCG GGG	GAA		1566 GCG GCA GGT  Ala Ala Gly
	CCG AT	r GGT 1		GGT	CCG	AGC GGC		CCG GGT	1620 CCG GAC GGC  Pro Asp Gly
1629 AAC AAA GGC			GTT GTT				ACC		1674 CCG TCT GGT
Asn Lys Gly	Glu Pr	o Gly 1	Val Val	Gly	Ala	Val Gly	Thr	Ala Gly	Pro Ser Gly
	CTG CO		GAA CGC		GCC		ATT		1728 GGC AAA GGT Gly Lys Gly
	GAA CO						AAC		1782 CGT GAC GGT Arg Asp Gly
	GCA CA		CG GTT				ccc		1836 GCG ACT GGT Ala Thr Gly
	GAA GC	r GGT (	CA GCG		ccc	GCG GC1	ccc	GCC GGC	1890 CCT CQC GGT  Pro Arg Gly
	GAA CG		GAA GIC		CCG		CCG		1944 TTT GCT GGC  Phe Ala Gly
							GAG		1998 GCC AAA GGC  Ala Lys Gly
2007		2016		2025		2034		2043	2052 OCG GCT GGT
Pro Lys Gly	Glu As	n Gly 1	Val Val	Gly	Pro	Thr Gly	Pro	Val Gly	Ala Ala Gly

	•	061		2	070		2	2079		2	980			2097			2106
CCG	GCT	GGC	CCĢ	TAA	GGT	CCG	CCG	CCT	CCC	GCA	GGC	AGC	CGT	GGC	GAT	GGT	GGC
Pro	Ala	Gly	Pro	Asn	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ser	Arg	Gly	Asp	Gly	Gly
	2	2115		2	2124		:	2133		2	2142			2151			2160
CCA	CCC	GGC	ATG	ACC	GGT	TTC	CCT	00C	GCG	GCC	GGT	ccc	ACC	GGC			GGT
Pro	Pro	Gly	Met	Thr	Gly	Phe	Pro	Gly	Ala	Ala	Gly	Arg	Thr	Gly	Pro	Pro	Gly
	:	2169		. :	2178			2187			2196	~~~		2205			2214
		GGC															
Pro	Ser	Gly	Ile	Ser	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Lys	Glu	Gly
		2223			2232	~>~	210	2241	~~		2250	CC4E		2259	CAA		2268
		GGC															
Leu	Arg	Gly	Pro	Arg	Gly	Asp	Gln	Gly	Pro	Val	Gly	Arg	Thr	Gly	Glu	Val	Gly
	~~~	2277 GGC	~~m	~~	2286	بلملمان		2295			2304 GGT			2313 GGT	GAA		2322 GGC
							~~-									~-~	~~~
Ala	Val	Gly	Pro								Gly	Pro	Ser	Gly	GIu		
100	~~	2331 GGC		~~·	2340	λOC	~	2349			2358 CCT			2367 CCT			2376
Thr	Ala	Gly	Pro	Pro	Gly	Thr	Pro	Gly	Pro	Gln	Gly	Leu	Leu	Gly	Ala	Pro	Gly
	~~~	2385			2394			2403						2421			2430
		CGC															
Ile	Leu	Gly	Leu	Pro	Gly	Ser	Arg	Gly	Glu	Arg	Gly	Leu	Pro	Gly	Val	Ala	Gly
~~	<b>⊘</b> ma	2439		~~	2448			2457			2466	~~		2475	~~		2484
GC1	GIA	GGC									901						GGI
Ala	Val	Gly	Glu	Pro	Gly	Pro	Leu	Gly	Ile	Ala	Gly	Pro	Pro	Gly	Ala	Arg	Gly
~~	~~	2493			2502			2511			2520	~~		2529			2538
		GGT				101			G11	AAC			CCT	GGT	GAA		GGC
Pro	Pro	Gly	Ala	Val	Gly	Ser	Pro	Gly	Val	Asn	Gly	Ala	Pro	Gly	Glu	Ala	Gly
~~	030	2547			2556			2565			2574	~~		2583			2592
	GAU	900	AAT			AAC	GAT				GGT	CGI	GAT.	GGT	CAG		GGT
Arg	Asp	Gly	Asn	Pro	Gly	Asn	Asp	Gly	Pro	Pro	Gly	Arg	Asp	Gly	Gln	Pro	Gly
CAC	222	2601 GGT		CCT	2610		ccc	2619			2628 CCT	œ		2637	CCC		2646
His	Lys	Gly	Glu	Arg	Gly	Tyr	Pro	Gly	Asn	. Ile	Gly	Pro	Val	Gly	Ala	Ala	Gly
ىتىت	~~	2655		C3C	2664		مىتى	2673			2682	***		2691	330		2700
		GGT															
Ala	Pro	Gly	Pro	His	Gly	Pro	Val	. Gly	Pro	Ala	Gly	ſŵs	His	Gly	Asn	Arg	Gly
CDA	<b>N</b> CC	2709 GGT		יראדי:	2718		وبلث	2727			2736	CCur	יואדי	2745	CC3	~	2754
Glu	Thi	Gly	Pro	Ser	Gly	Pro	. Val	. Gly	Pro	Ala	Gly	Ala	Val	Gly	Pro	Arg	Gly

		2763			2712			2781			2790			.:799			2808
CCC															GAΛ		
Pro	Ser	Gly	Pro	Gln	Gly	Ile	Arg	Gly	Asp	Lys	Gly	Glu	Pro	Gly	Glu	Lys	Gly
	:	2817			2826			2835			2844		;	2853		:	2862
CCCG	CGT	CCT	CTG	CCC	GGC	CIT	AAG	GGC	CAC	AAC	CCT	CTG	CAA	GGT	CTG	CCC	GGT
Pro	Arg	Gly	Leu	Pro	Gly	Leu	Lys	Gly	His	Asn	Gly	Leu	Gln	Gly	Leu	Pro	Gly
		2071			2000			2000			2000			2007			
		581T	~~~	030	2000	Cam	CAC	2007 CCIII	~	~~	2070	m~	~~	7907	~~~	~~~	2916
AIC	·GCG	GGT	CAC	CAC	COL	GAT	CAL	991	CLI	CCG	GGT	ncc	GIT	GG1	CCC	GCC.	GGT
					01	200	01-	01	212	ò	01	5000	7/- 1	03			
ITE	ALG	GTĀ	HIS	HIS	GIA	ASD	GIII	GIŞ	WTG	PIO	GIA	zer	var	GIĀ	Pro	Ата	GIĀ
		2925		:	2934			2943			2952		:	2961		:	2970
CCG	CGT	GGC												GGC	CGT		
Pro	Arg	Gly	Pro	Ala	Gly	Pro	Ser	Gly	Pro	Ala	Gly	Lys	Asp	Gly	Arg	Thr	Gly
		2020			2000			2007			2005		_				_
C3.C		2979		~~	4900 ~~	~~	-	2331	3.000		3006			3015		3	3024
CAL		CCI	ACG	GIG	GGI	CCG	GCC	GGC	AIT	CGC	GGT	CCG	CAA	GGT	CAC	CAG	GGT
His	Pro	Glv	Thr	Val	Glv	Pro	Ala	Glv	Ile	Arro	Glv	Pro	Gln	Gly	His	C12	Clv.
																	-
		3033			3042			3051			3060		3	3069		3	1078
$\infty$	GCG	CCT	$\infty$	CCC	GGT	$\infty$	$\infty$	CCT	$\infty$	CCG	CCT	$\infty$	CCCG	GGT	GTT	AGC	CCT
Pro	Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Val	Ser	Gly
	-	1027			RNGE			2105									
GGC	GGT	TAT	GAT	datal.	TODO TODO	TAT	GAC.	1200 1	CMT	mm	MVW.	~~	000	٠.			
								301	GMI	TIC	TAT	CGI,	GCG	3'			
Gly	Gly	Tyr	Asp	Phe	Gly	Tyr	Asp	Gly	Asp	Phe	Tyr	Arg	Ala				

FIG. 50E

 $\overline{\epsilon_{co}R_1}$  start Oigo N1–1 ( $lpha_2$ ) 5-66aattcatgccatgatgatgatgaccaaaggcgtcgccttggcccccaatggccctcatgggccccccgcggccca-3'

BsrFl stop Hind III

3'-CCGGGCGCGCCGCGGGGCCCACGTCGACCGCGGGGTCCGGGGTCCCGGGACGCCCGAATTATTCGAACCC-5'

Oligo N1-2 ( $\alpha_2$ )

EcoR1 BSRF1 Oligo N1-3 ( $\alpha_2$ )

S'-GGAATICGCCGGTGAGCCGGGTGAACCGGGCCAAACGGGTCCGGCGGCGCGCGTGGTCCAGCGGGCCGCCCTGGCAAGGCG-3'

3'-CCGGGCGGACCGTTCCGCCACTTCTACCGGTGGGACCGTTTGGCCCGGGGGCCCACTCGCACCGCATCACATATTCGAACCC-5'

Oligo N1-4  $(\alpha_2)$ 

FIG. 51

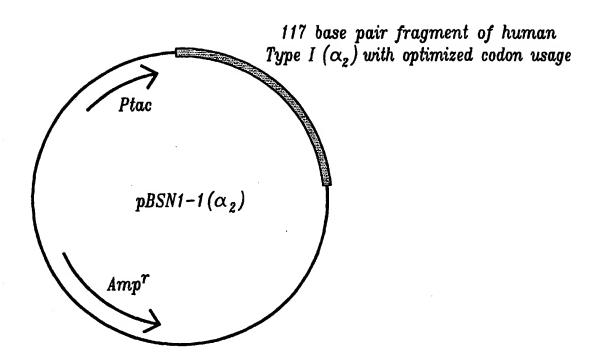


FIG. 52

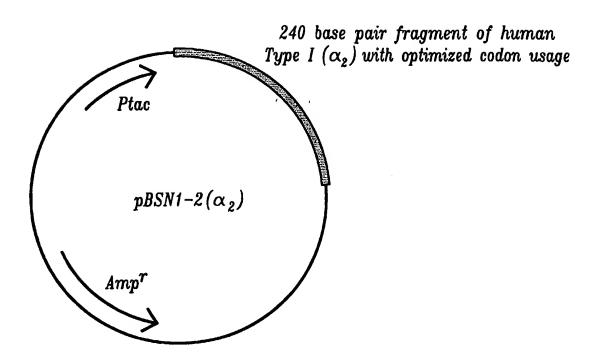


FIG. 53

		9				18			27			36			45			54
5'	CAG	TAT	GAT	GGC	AAA	GGC	GTC	GGC	CTC	GCC	$\infty$	GCC	CCA	ATG	GGC	CTC	ATG	GGC
	Gln	Tyr	Asp	Gly	Lys	Gly	Val	Gly	Leu	Gly	Pro	Gly	Pro	Met	Gly	Leu	Met	Gly
			63			72			81	•	•	90			99			100
	~~	~~			~~		CCN	~		~~	~~			~				108
	us	us	GGC	u.	u.	GGI	GCA.	CCI	GGC	مبد	CA	GG	فللا	CAA	GGI	TIC	CAG	GGC
												_						
	Pro	Arg	Gly	Pro	Pro	Gly	ALa	Ala	Gly	Ala	Pro	Gly	Pro	Gln	Gly	Phe	Gln	Gly
			117			126			135			144			153			1.00
	$\sim$ r	ccc	GGT	GNG	m						300			~~.	777			162
	~							uu.	GGC.	CM	MUS	GGI	u	GCA	GGT	GCA	CGI	GGT
	D	31-	C1	Δ1	7	~l	C)	D	~									
	PIO	ALG	Gly	GTI	PIQ	GTĀ	GLU	PEO	CTA	GLR	The	GŢĀ	Pro	Ala	Gly	Ala	Arg	Gly
			171			180			189			198			207			21.5
	CCA	GCG	GGC	$\infty$	CT	GCC	AAG	CCG	GGT	(A)	ωт	220	CAC	~	201		~~~	216
											GET	نانان	CHL	α.i	GGC	AAA	CC;	GGC
	Pm	Δla	Glu	Pm	Pro	Clv	Tare	21-	C1	Cl	<u></u>	~~						
			Gly		***	GLy.	mys	ALC.	GLY	GLU	ASD	GrÅ	HLS	Pro	Gly	Lys	Pro	Gly
			225			234												
	$\infty$	$\infty$	<b>GCT</b>	GAG	ŒT	GCC	GTA	GTG										
			<u></u>															
	Arg	Pro	Gly	Glu	Arg	Gly	Val	Val										

FIG. 54

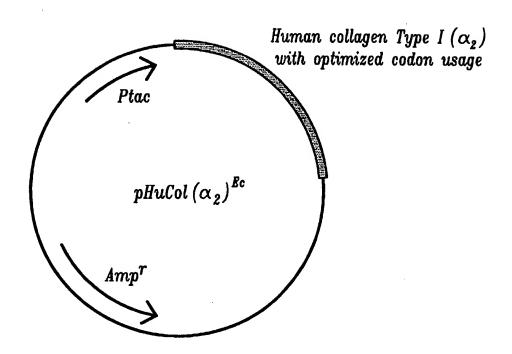


FIG. 55

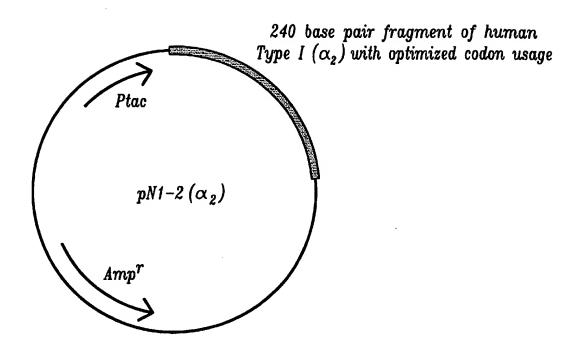


FIG. 56

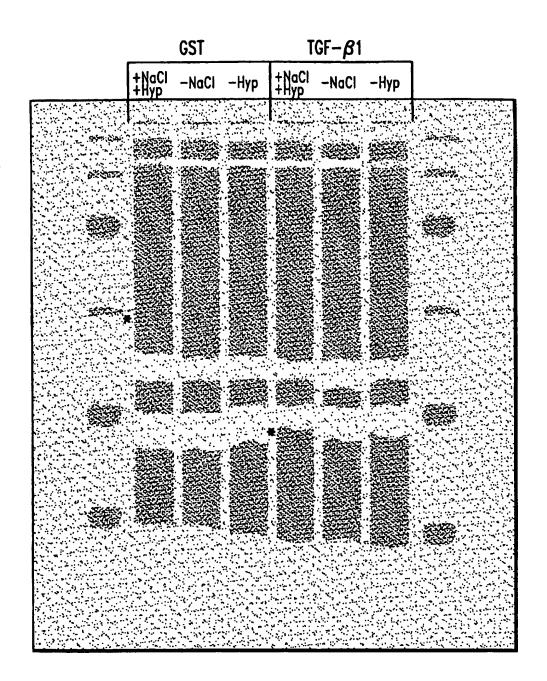


FIG. 57

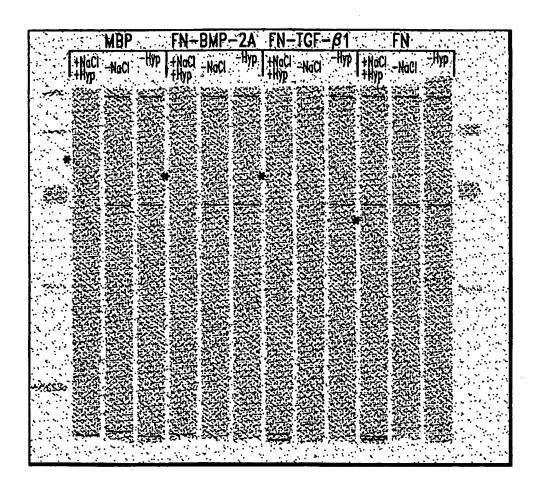


FIG. 58

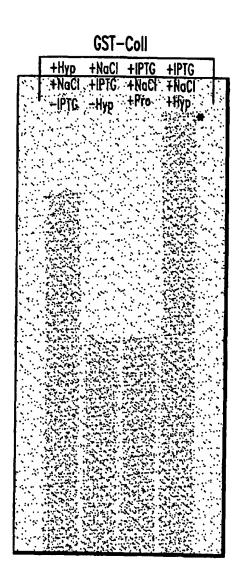


FIG. 59

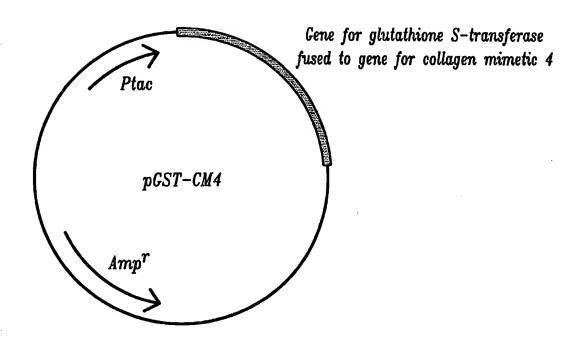
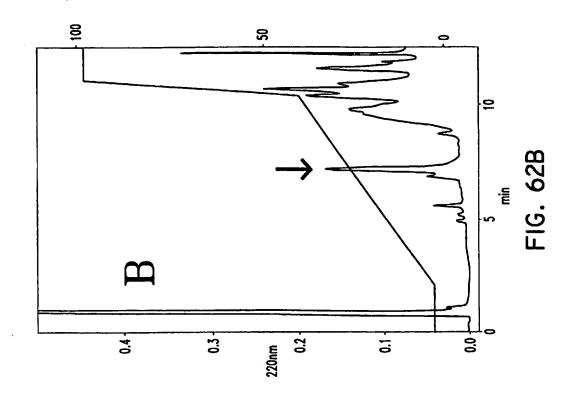
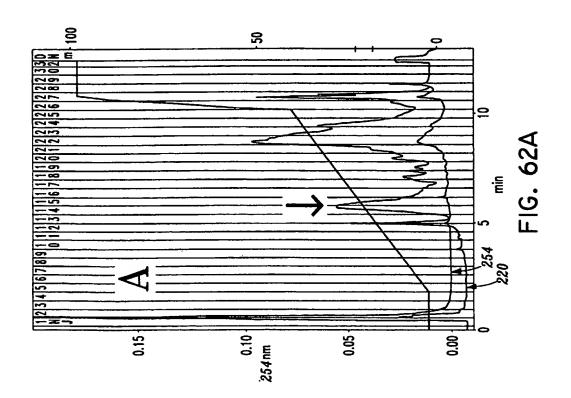


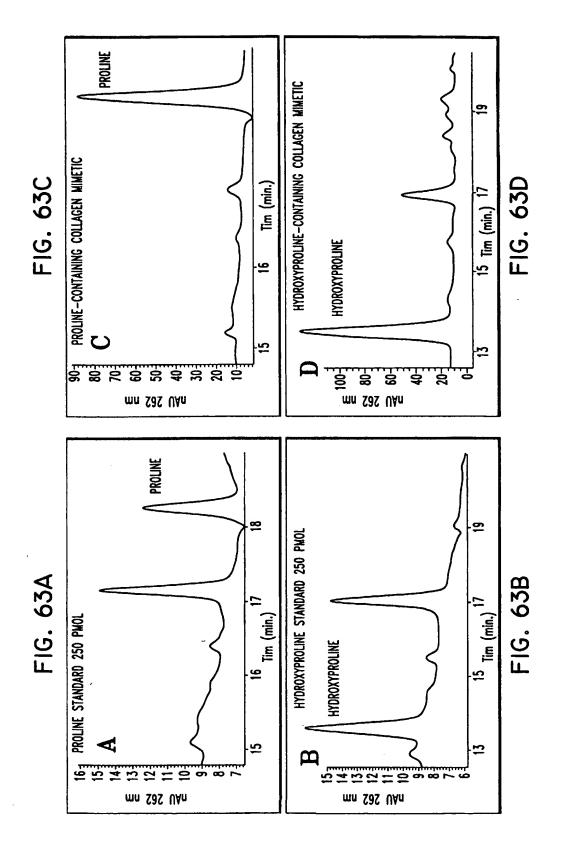
FIG. 60

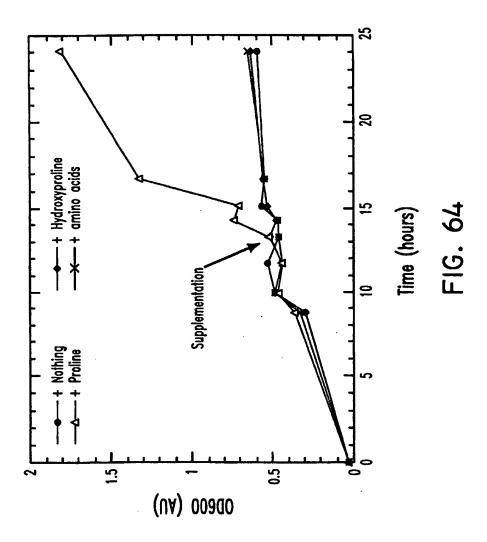
			9			18			27			36			45			54
5'	ATG	GGG	CTC	GCT	GGC	CCA	CCG	GGC	GAA	CCG	GGT	CCG	CCA	GGC	CCG	AAA	GGT	CCG
	M	G	L	Α	G	P	P	G	E	P	G	P <sub>.</sub>	P	G	P	K	G	P
		•	63									90						108
	CGT	GCC	GAT	AGC	GGG	CLC	GCT	GGC	CCA	CCC	GGC	GAA	CCG	GGT	CCG	CCA	GGC	CCG
		- <del>-</del>																
	R	G	D	S	G	L	A	G	P	P	G	E	. P	G	P	P	G	P
			117			126			135			144			153			162
	444	GGT										CCA						
	ĸ	G	P	R	G	D	S	G	L	A	G	p.	P	G	E	P	G	P
		:																
		;	171			180			189			198			207			216
	CCA	ĢGC	CCG	AAA	GGT	CCG	CGT	GGC	GAT	AGC	GGG	CTC	CCT	GGC	CCA	CCG	GGÇ	GAA
		1																
	₽	\; G	P	K	G	₽.	R	G	Đ	S	G	L	A	G	P	P	G	E
		÷	225			234			243			252			261		٠,,	270
	CCG	ĠGT	CCG	CCA	GGC	CCG	AAA	GGT	CCG	CGT	GGC	GAT	AGC	GGG	CTC	CCG	GGC	GAT
	P	G	P	P	G	P	K	G	P	R	G	D	S	G	L	P	G	D
		:																
	TCC	TAA	31															
			•															•
	s	*	•															

FIG. 61









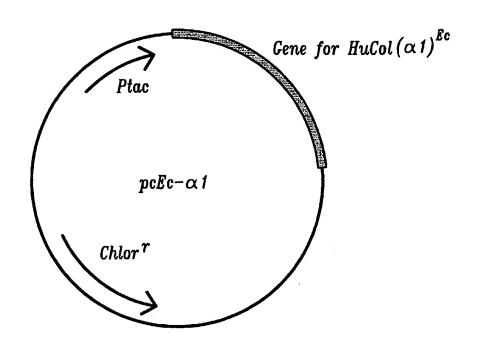


FIG. 65

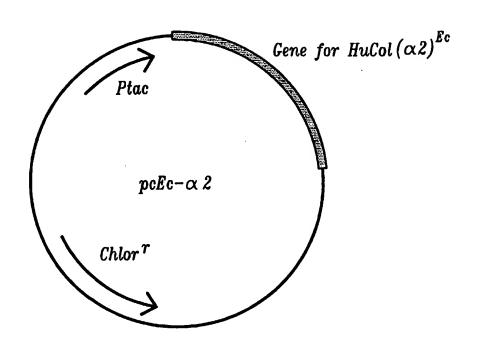


FIG. 66

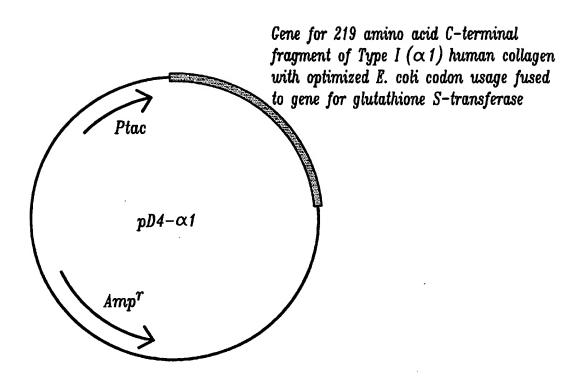


FIG. 67

Gene for 207 amino acid C-terminal fragment of Type I ( $\alpha$ 2) human collagen with optimized E. coli codon usage fused to gene for glutathione S-transferase  $pD4-\alpha 2$ 

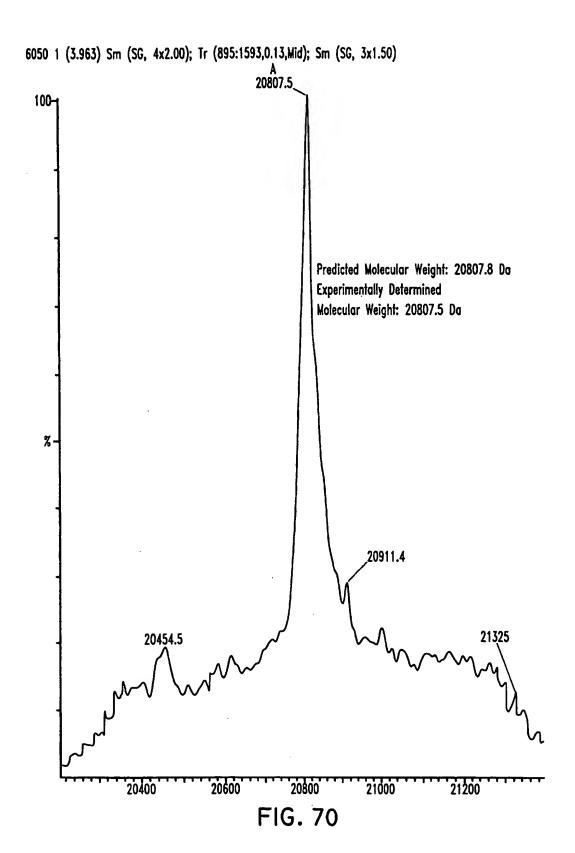
FIG. 68

Protein Sequence of the First 13 amino acids of D4-α1. Predicted From the DNA Sequence: H<sub>2</sub>N-Gly-Pro-Pro-Gly-Leu-Ala-Gly-Pro-Pro-Gly-Glu-Ser-Gly

Experimentally-Determined Protein Sequence of the First 13 amino acids of D4-α1:

H<sub>2</sub>N-Gly-Hyp-Hyp-Gly-Leu-Ala-Gly-Hyp-Hyp-Gly-Glu-Ser-Gly

FIG. 69



		9			18		ccc	27	ccc	CCD	36	<b>&gt;</b> CC	CCT	45	CAA	· •	54 GCG	ccc	com
000	C) }	69	NCC	CC 1	78	ccc	CAC	87 (CT)	λCC	ccc	96	.ccc	444	105	СУТ	ርርጥ	114 GGT	CAA	) CC
ιις	GAA																	ann	ACC
ייטט	cna	129	cee	(°C)	130	CCT	GCA	147 CCG	GGC	GCG	156 ccc	GGT	GCC	165 CCA	GGC	ccs	174 GTG	ccc	ccc
000																			
GC G	GGC	133	VCC	GGT	GAT	CGT	GGT	GAG	RCC	GGT	CCC	GCG	GGC	CCC	GCC	GCT	CCC	GTG	GGC
		244			253			267			276			285			294		
CCA	GCG	CGC	GCC	CGT	GGC	CCG	GCC	GCT	CCG	CAG	GGC	CCG	CGG	GGT	GAC	AAA	GGT	GAA	ACG
		309			313			327			336			345			354 CAG		
GGC	GAA	CAG	CCC	GA)	CCT	GGC	ATT.	AAA	GGC	CAC	CGT	GGC	TTC	AGC	GGC	CTG	CAG	CCT	CCA
0,0		369	000	000	378	ccc	~~	387	C>C		396		<b></b>	405			414 CCG		
CCG	GCC																	GCG	GGC
CCA	CGC	429 GGT	CCG	င်ငဒ	438 GGC	AGC	CCG	447 GGC	GCG	ccc	456 CGC	AAA	GAC	465 GGT	CTC	<b>726</b>	474 GGT	CTYC:	ccc
•••																		CiG	CCG
CCC	CCG	ATC	GGC	CCG	CCG	GGC	CCA	CGC	CGC	CGC	YCC 21'0	GGT	GAT	CCC	GGT	CCG	GTG	GGT	ccc
		549			558			567			576			585			504		
CCC	GGC	cce	CCG	GGC	CCG	CCA	GGC	CCG	CCG	GGA	CCG	CCG	AGC	GCG	GGT	TTC	GAC	TTC	AGC
		609			61.8			627			636		•	645			654		
TTC	CIC	CCG	CAG	CCG	CCC	ÇAG	GAG	AAA	GCG	CAC	GAC	GGC	CCI	ccc	TAC	TAC	CCT	GCG	TAA
		669			673			687			696			705			714		
		• • •	• • •	• • •	• • •	• • •	• • •	•••	•••	•••	• • •	•••	• • •	• • •			• • •		• • •

FIG. 71

60	50	40	30	20	10
APGAPGPVGP	GPAGPPGAPG	PGAKGDRGET	REGSPOREGS	GESGREGAPG	MGPFGLAGPP
120	. 110	100	90	30	70
HRGFSGLQGP	GEOGORGIKG	CGPRGDKGET	PAGARGPAGP	GPAGPAGPVG	AGKSGDRGET
180	170	160	150	140	130
RTGDAGPVGP	GPIGPPGPRG	PGKDGLNGLP	PRGPPGSAGA	GI'SGASGPAG	PGPPGSPCEQ
. 240	230	220	210	200	190
		HEGGEYYRA*	FLPOPFQEKA	GPPEAGEDES	PGPPGPPGPP

FIG., 72

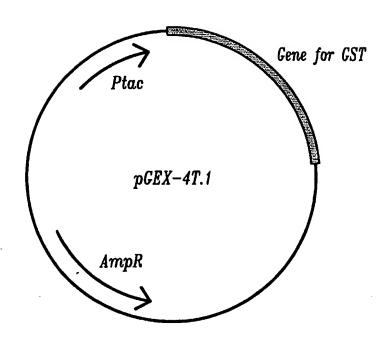


FIG. 73

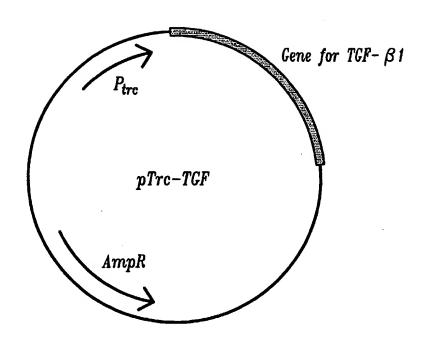


FIG. 74

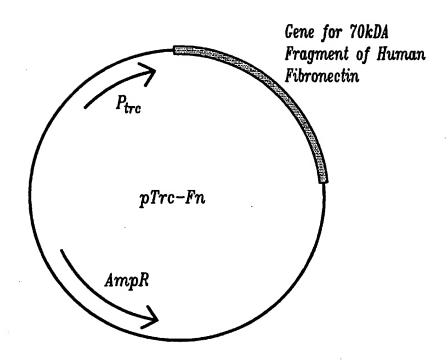


FIG. 75

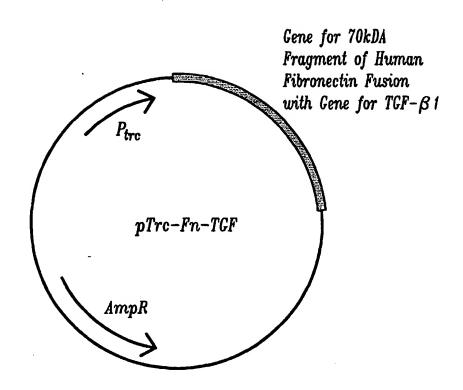


FIG. 76

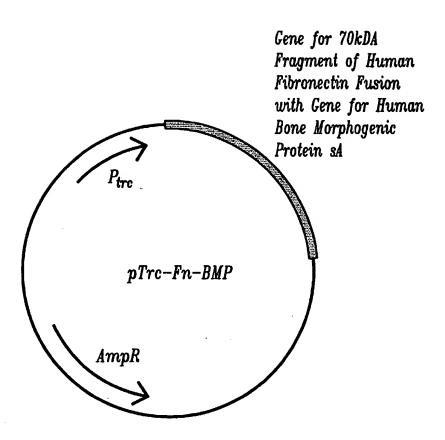


FIG. 77

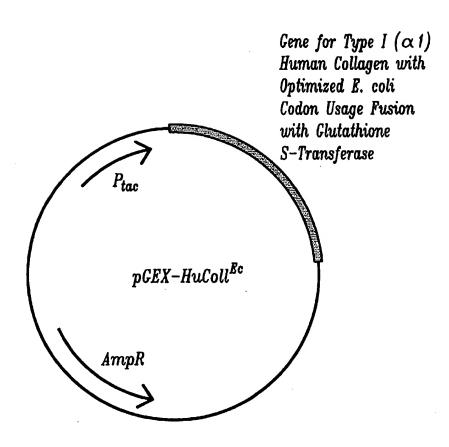


FIG. 78

FIG. 79

FIG. 80

#### Oligo N4-1

5'GGAATTCTCCCATGGGCCGGCCGGGTCTGGCGGGCCCTCCGGGTGAAAGCGGTCGTGA AGGCGCGCGGGTGCCGAAGGCAGCCCAGGCCGCGAC

#### Oligo N4-2

3'CTTCCGTCGGGTCCGGCGCTGCCATCGGGCCCCCGGTTTCCCCTAGCACCACTTTGGCC GGGCCGCCCGGGGGGCCCACGTGGCATTATTCGAACCC

### Oligo N4-3

5'GGAATTCGGTGCACCGGGCGCGCGGGCGGGCAAA AGCGGTGATCGTGGCGAGACCGGTCCGGCGGGC

### Oligo N4-4

3'CTCTGGCCAGGCCGGGCCGGGCCAGGCCACCGGGTCGCCGGGCACCGGGCC GGCCAGGCGTCCCGGGCGCCATTATTCGAACCC

# This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

# **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

| BLACK BORDERS
| IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
| FADED TEXT OR DRAWING
| BLURRED OR ILLEGIBLE TEXT OR DRAWING
| SKEWED/SLANTED IMAGES
| COLOR OR BLACK AND WHITE PHOTOGRAPHS
| GRAY SCALE DOCUMENTS
| LINES OR MARKS ON ORIGINAL DOCUMENT
| REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

IMAGES ARE BEST AVAILABLE COPY.

OTHER:

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.